

Pros and cons of centralization

A proposal by Germany's science minister to concentrate power over much of the country's best research is less monstrous than some are suggesting, but still leaves plenty to worry about.

Few countries have as much power vested in regional states as Germany. This federalism was a response to the centralization of power and 'Germanization' of education, science and art that occurred under National Socialism in the 1930s. In science and education, as in many other areas, this federalism has created a complicated system in which the responsibilities of the federal government and the 16 *Länder* (state) governments are intricately interwoven — too complicated to function efficiently. In a bid to revitalize the crisis-ridden country, the German government now plans to simplify matters.

Germany's research minister, Edelgard Bulmahn, may have sensed a political opportunity. Lacking profile in the cabinet, she needs a success. She wants to centralize funding of the Max Planck Society and the DFG, Germany's main research funding agency, while giving the *Länder* full financial responsibility both for university infrastructure and for non-university research carried out under the umbrella of the Leibniz Association. Her ideas have received a hostile reception from researchers (see page 790 in this issue).

But Germany's research managers are dismissing the idea too hastily. They say that centrally funded basic research could become overly biased towards political and strategic objectives. However, the world is full of successful examples of such centralized structures, in which funding agencies are nevertheless protected from direct political control. Indeed, a streamlining of the federal research system's cumbersome decision-making procedures is long overdue.

Giving the *Länder* full responsibility for universities, for example, would allow them to find solutions that better match local needs and requirements. This would be a vast improvement over the existing system, in which each new X-ray machine or building project

that is backed by the *Länder* must also be approved at a federal level.

But maintaining and renewing university infrastructure is a costly and thankless task. The federal government's joint responsibility has a regulating effect, keeping in check the disparities in the quality of universities' research. Germany's science system would benefit if the universities were forced to rethink their strengths and weaknesses, and develop more distinctive profiles. The introduction of tuition fees could also support these efforts.

There are pitfalls to be avoided, however. True, the *Länder* have a vital interest in maintaining universities, just as they have in building and maintaining schools, regional roads and parks. But they would be less enthusiastic about the rest of the legacy. In particular, most Leibniz institutes have little or no regional linkage, so why should the state of Brandenburg, for example, be interested in funding the Leibniz institutes of astrophysics and climate research in its capital, Potsdam? What is to stop it closing these laboratories?

The Leibniz institutes are the least known and most diverse pillar of German research. Bulmahn's proposals are in effect a policy of abandonment that risks eroding Germany's arduously restructured research landscape. Perhaps the best Leibniz laboratories can find a new home at government-funded national research centres, or in the Fraunhofer Society for applied research?

Although more sensible than research managers would like to admit, Bulmahn's proposed reforms have put her out on a limb. In their current form, they might result in a cutback of overall research expenditure, the very opposite of the German government's declared goal. But with more thought and much fine-tuning, the reforms could be good for German research. ■

Silence of the neuroengineers

Researchers funded by a defence agency should stop skirting the ethical issues involved.

Neuroengineering is in its infancy, but already provides plenty of food for thought. Ethicists can cite a slew of projects that they have studied while trying to come to grips with the new technology. How will it affect human identity, for example? And could it one day be used to control thought processes?

Not so long ago, the idea of connecting a tiny machine to the nerves of the heart had ethicists worried — would it make us in some way less human? Today, over half a million pacemakers are implanted annually without any soul searching. As with many new technologies, the reality turned out to be less disturbing than some had speculated.

Integrating machines with the brain will be more ethically complex. One research group has shown how to control the movement of a rat by sending signals to electrodes implanted in its brain. Others have taught a monkey to control a robot arm using signals taken directly from its brain (see page 796). The scientists involved are happy to speculate about their work; some say that the robot-arm experiments, for instance, could lead to a new generation of prosthetic limbs.

But the researchers should perhaps spend more time pondering the intentions of the people who fund their work. A significant amount

of US neuroengineering research is funded by the military through the US Defense Advanced Research Projects Agency (DARPA). The agency wants to create systems that could relay messages, such as images and sounds, between human brains and machines, or even from human to human.

In the long term, military personnel could receive commands via electrodes implanted in their brains. They could also be wired directly into the equipment they control. Do neuroengineers support these goals? Their research could make it happen, so they have a duty to discuss their opinions, and to answer questions from those who object to the development of such technologies. Yet when *Nature* talked to DARPA-funded neuroscientists, many were reluctant to debate the potential military uses of the technology, saying that the agency's goal of brain-machine interfaces was still many years off.

The agency's goal may indeed be a distant one. But like all new potential technologies, it is worth discussing the consequences now. Simply taking DARPA's money, and citing possible medical benefits, is not enough. The discussions may never achieve a consensus, but will achieve a better quality and balance with researchers' engagement. ■





Staple diet
Funding reprieve for traditional approach to breeding wheat
p790



All together now
SARS encourages international collaborations
p791



Delta blues
Ecologists fear for the future of the Colorado River
p793



Green shoots
Physics lab lends Hulk a helping hand
p794

US army attacked over published patent for 'bioweapons grenade'

Erika Check, Washington

Critics are charging that a patent granted to the US army raises troubling questions about the way the military handles information on biological and chemical weapons.

The patent, issued in February, is for a "rifle-launched non-lethal cargo dispenser" that can be filled with "smoke, crowd control agents, biological agents, chemical agents, obscurants, marking agents, dyes and inks, chaffs and flakes" (N. Gonzalez *et al.* US patent 6,523,478; 2003). It goes on to describe in detail how to make the projectile.

Greg Aharonian, a patent consultant based in San Francisco, says that the patent constitutes step-by-step instructions for making a lethal weapon. "It is hypocritical to complain about countries developing biological and chemical weapons when we are openly educating them on how to do so," he says.

Other observers doubt whether the patent is a real security risk. "I don't think this would be a particularly effective munition for delivering biological agents, so I don't see its publication as a security issue," says Mark Wheelis, a microbiologist at the University of California, Davis, who advises the Federation of American Scientists on bioweapons issues.

But Wheelis and others are still concerned that the patent's claims made it through the Department of Defense's internal review process. The Pentagon reviews its inventions to see if publishing patents on them could threaten national security, and it seals potentially threatening ones behind secrecy orders.

The critics say that the reviewers should have known that the United States is banned from producing biological weapons, both by the Biological Weapons Convention and by federal law. "The army needs to assure the country and the international community that everybody who reviews these projects is familiar with the obligations of international and domestic law," says Wheelis.

Miguel Morales, a spokesman for the army's Edgewood Chemical Biological Center in Aberdeen, Maryland, says that the army has no intention of using the patented grenade as a biological or chemical weapon. He says that reference to biological and



The US is destroying its chemical weapons stockpile, but is its research breaking international rules?

chemical agents was inserted by a lawyer who was trying to cover as much ground as possible with the patent — a common practice in intellectual property filings.

"We're not violating any treaty, and we're going to be deleting those words in the patent to ensure that there is no further misunderstanding," Morales says.

But the critics contend that deleting the troublesome words from the patent does not

answer the larger question of why they were written into the application in the first place.

"To have that description openly in the application reveals some major problems in the way the United States thinks about that particular weapons system — and how it's thinking about the larger question of biological and chemical agents," says David Fidler, an international lawyer at Indiana University in Bloomington. ■

Whaling group backs conservation

Charlotte Westney

The International Whaling Commission (IWC) has voted in favour of a fundamental shift in its remit, formally recognizing for the first time that conservation is part of its job.

The decision, carried by 25 votes to 20, was made on 16 June at the commission's 55th annual meeting, held in Berlin.

Members agreed to set up a conservation committee within the IWC, which will examine the effect of whaling and other activities, such as commercial fishing, on a broad range of cetacean species. Until now, the IWC primarily considered how to manage the whaling industry.

But the vote leaves a deep rift in the commission. Countries that support commercial whaling, such as Japan, Norway

and Iceland, all said that they will not participate in or fund the new committee. Minoru Morimoto, Japan's commissioner at the IWC, said that the decision to protect all whales irrespective of conservation status undermined more than ten years of work by the IWC's scientific committee into sustainable commercial whaling.

The vote was enthusiastically received by more than 40 conservation groups that supported it. They hope that the committee will address issues such as the 300,000 whales, dolphins and porpoises that are killed each year when they become entangled in fishing nets. "This is a historic day for cetacean conservation," says Susan Lieberman, who led a delegation at the IWC talks sent by the environmental organization the WWF. ■

Traditional wheat breeder fights for funding

Rex Dalton, San Diego

A prominent wheat breeder at a major US public university has narrowly retained funding for his work, after fighting off what supporters describe as a concerted effort by commercial seed suppliers to get it cut.

Stephen Jones, a plant geneticist at Washington State University (WSU) in Pullman, refuses to participate in industry-funded projects, and has spoken out publicly against what he sees as corporate efforts to gain control of wheat farming through biotechnology.

Many wheat farmers use their own seed, held over from the previous year's crop (a practice largely dropped by farmers of other crops, including corn and soya beans). Jones uses traditional breeding methods to produce more disease-resistant varieties of winter wheat that can be used in this way.

In March, some farmers—encouraged by local seed suppliers, according to several growers and officials—started a campaign to get Jones's core funding of some \$200,000 a year cut. They wanted money spent on projects that would involve commercial partners and develop crops containing patented genetic mutations. If a seed contains a patented trait it can't be legally replanted.

Jones claims that some farmers were misled. "They didn't realize that they would be destroying the public winter-wheat breeding programme at WSU," he says.

After weeks of political manoeuvring, the Washington Wheat Commission, which helps to fund WSU research programmes, voted last month to underwrite Jones's studies for the next year. The board of the Washington Association of Wheat Growers, a 3,000-member group that strongly influences the commission, voted six to five in favour of Jones's research.

Some growers, who didn't want to be identified, said the fight reflected anxieties among farmers, who want to produce wheat that is competitively priced but that will be saleable in parts of the world where transgenic varieties aren't acceptable to consumers.

Many farmers remain sceptical about the transgenic wheat being prepared for the market by large suppliers such as Monsanto. They want competitive, non-transgenic varieties based on patented genetic mutations. "Wheat growers are desperate to make ends meet," says grower James Moore, a former Wheat Commission member from Kahlottus in eastern Washington. "But the idea of cutting the wheat-breeding research programme is beyond belief. It's the only thing that will save the wheat industry."

Ralph Cavalieri, associate dean of the WSU College of Agriculture and Home Economics where Jones works, says Jones will continue his research, alongside a new industry-assisted programme. The pro-



Gold reserve: many farmers save seed to replant.

gramme will involve a private company, Pullman-based Northwest Plant Breeding, and will seek to develop a new winter wheat containing a herbicide-resistant trait called Clearfield, owned by seed firm BASF.

But Cavalieri remains worried about the future of publicly funded plant breeding (see *Nature* 421, 568–570; 2003). "As public support of universities such as ours diminishes, our ability to provide research in the public domain declines," he says. "I have concerns about the sustainability of these public breeding programmes."

German reform plan arouses fears for autonomy

Quirin Schiermeier, Munich

German scientists and research managers are up in arms over a plan to overhaul laboratory and university funding.

The plan, to shift responsibilities between federal government and Germany's 16 *Länder* (states), could threaten the autonomy of basic research institutions, critics claim.

Germany's postwar constitution,

designed to avoid the concentration of power, split many responsibilities between federal government and *Länder*. Support of research laboratories, agencies and universities is split evenly between the two.

But under a government-wide reform programme set out last month, federal government would take sole responsibility for the 80 labs and institutes operated by the Max Planck Society (MPS) and for the DFG, Germany's main granting agency.

The *Länder* would maintain universities and 80 research institutes, museums and libraries operated by the Leibniz Association. Research minister Edelgard Bulmahn says that centralized financing would streamline decision-making.

And some observers recognize the need for change. "Germany's extreme federalism can indeed be obstructive," says Ulrich Herbert, a historian at the University of Freiburg, and a member of the Wissenschaftsrat, Germany's independent science council. "Untangling responsibilities could be a good thing," he adds, "but not if the federal government takes the pick of the

bunch, and gives the *Länder* what's left."

MPS president Peter Gruss opposes the change, saying: "Before you replace a tried and tested system you should provide proof" of the substitute's superiority. Ernst-Ludwig Winnacker, president of the DFG, says that the "checks and balances" of the federal system underpin researchers' freedom from political interference.

The government and *Länder* must agree by the autumn, when Chancellor Gerhard Schröder wants to launch the reforms. He will start talks with state prime ministers later this month, and amendments to the proposal are likely.

Meanwhile, Germany's budget crisis, fuelling the reforms, is also hitting research directly. At the MPS's annual assembly in Hamburg on 5 June, Gruss announced a list of 13 research departments that would be closed by 2007, including departments at the MPS institutes of physics and quantum optics in Munich, experimental medicine in Göttingen, and radioastronomy in Bonn. The planned expansion of nine other institutes will also be dropped.



China and Germany join forces over SARS

Alison Abbott, Munich

When Jiang Hua-Liang and Rolf Hilgenfeld ran into each other in the empty corridors of a Beijing hotel earlier this month, they immediately discovered how much they had to talk about.

Now they are hoping to work together to study the virus believed to be behind severe acute respiratory syndrome (SARS). The planned project is just one example of a host of unexpected international alliances that have sprung up in the wake of the mystery disease.

Jiang heads the Drug Discovery and Design Center at the Shanghai Institute of Materia Medica, and Hilgenfeld is a structural biologist at the University of Lübeck in Germany. The pair were participating in a workshop in the Chinese capital on 3–4 June, organized by the Sino-German Centre for Science Promotion to encourage collaboration between SARS researchers.

The German delegation of 14 virologists was the first of its type to arrive in China since the SARS epidemic began, and it arrived to find that fear of the disease had emptied Beijing's streets of residents and tourists.

Hilgenfeld is an expert in coronaviruses, the family to which the SARS virus belongs, and has detailed crystal structures of a key SARS virus protein. Jiang uses supercomputers for the mass screening of drug candidates — identifying molecules that fit snugly into computerized three-dimensional models of proteins. The researchers realized how efficiently they could work together to develop new leads for anti-SARS drugs.

The idea is one of a dozen or so that over the next two months will be turned into formal funding proposals for the Sino-German Centre, which is jointly financed to the tune of €2 million (US\$2.4 million) annually by research agencies in Germany and China.

The German visitors say that they were taken aback to see how far and fast Chinese scientists have come since turning their attention to the SARS virus a couple of months ago.

China has had thousands of SARS cases and, consequently, has a wealth of patient material and clinical experience that the German scientists — who only have isolates from two clinical cases that passed through Frankfurt airport in March — want to access.

"Benefits of the collaborations cut both ways," says Hans-Dieter Klenk, head of Marburg University's Institute of Virology, who led the delegation. "Germany has good basic research experience with coronaviruses." It also has important tools such as a new-generation polymerase chain reaction (PCR) test — a method used to amplify selected regions of DNA — which is waiting to be tested in patients.

But the German researchers say that the Chinese scientists are strongly possessive of their virus isolates. They insist that it would be too dangerous to send them to Germany, and that any collaborative work should therefore be done in China. "They are understandably concerned not to lose control of their resources," says Klenk.

The Germans add that they are happy to work in China on clinical questions such as the immune response to the virus, which will benefit from the hundreds of different isolates the Chinese have, scores of which have already been fully sequenced. Christian Drosten, a clinical virologist at the Bernhard Nocht Institute for Tropical Medicine in Hamburg, says that he is also optimistic about potential collaborations to put his PCR test through its paces, but says that he realizes it may not be so easy in practice.

Some of the potential collaborations will address more basic research questions, such as understanding how the virus regulates its own gene expression, which can be done without working on the entire virus.

The visitors were also struck by the high



Common goal: Jiang Hua-Liang (centre) explores collaboration with German researchers in Beijing.

laboratory standards, and the commitment of the Chinese scientists. "I was really surprised to see so much investment in high-tech laboratory equipment," says Hilgenfeld, who paid a late-night visit to Tsinghua University in Beijing and found droves of researchers still beaver away at 11 o'clock. "You would never see that in Germany," he says. ■

Clinical data marshalled to treat HIV

Tom Clarke, London

A database is set to help doctors treat HIV. It attempts to identify the best drug combinations to keep a patient's infection in check, on the basis of genetic sequences of drug-resistant strains of the virus.

The international collaboration that is developing the system — the HIV Resistance Response Database Initiative (RDI) — unveiled its latest results on 12 June at the 12th International HIV Drug Resistance Workshop in Los Cabos, Mexico. And although the system isn't the first to deduce information about the drug resistance of particular HIV strains, it is the first to attempt to do so on such a scale, and to be based on clinical results.

"We're aiming to produce tools that can predict treatment response in actual patients," says Brendan Larder, a virologist who chairs the initiative's core group. The system could eventually be available to all physicians online, enabling them to use a patient's HIV gene sequence to help to predict what drug treatment options would be most successful.

Rapidly mutating HIV strains readily develop resistance to drugs. Up to half of all patients living with HIV must switch drug combinations at least once to control their viral load. Studies have identified about 200 mutations associated with

resistance to the 16 drugs commonly given to control the virus.

Doctors already use computer models to establish the optimal combination of two or three drugs that should be given to treat a particular, resistant strain of HIV. But that is an inexact business, given that a patient's virus may harbour as many as 40 drug-resistance mutations. Also, input to the existing computer models usually derives from the ability of different strains to survive in test-tube experiments — which may not reflect the outcomes in patients.

"What's missing is a link to clinical response," says Joep Lange, director of the National AIDS Therapy Evaluation Center in Amsterdam. "At the moment, a lot of mistakes are being made." Because the RDI software 'learns' from real patient data, it could surmount these problems, he says.

The latest test of the software showed that the system could correctly predict which drug combinations would bring HIV under control, and by how much, 78% of the time.

This isn't reliable enough for clinical use, Larder admits. But the latest results are from the 750 cases in the original database. And data from about 4,000 patients at AIDS treatment centres in Europe and North America are currently being added. Once that happens, "we hope to achieve 90% accuracy," says Larder. ■

British panel bans use of antidepressant to treat children

Alison Abbott

UK regulators have ruled that a top-selling antidepressant should not be prescribed to children or adolescents who are suffering from depression.

The drug, paroxetine, is an ineffective treatment for major depression in this age group and could possibly increase suicidal tendencies, the Committee on Safety of Medicines (CSM) said on 10 June.

Paroxetine, a member of the same class of drug as Prozac, is widely used to treat depression and other nervous disorders. It is not licensed for use in children but can be prescribed by doctors on an individual basis — about 8,000 patients aged under 18 have been given the drug in Britain in the past year.

The CSM's ruling followed a review of fresh clinical data on 1,200 children treated with paroxetine for depression, social anxiety and obsessive-compulsive disorder. The information was provided by UK-based drug firm GlaxoSmithKline (GSK), which markets the drug under the name of Seroxat in Britain and as Paxil in the United States.

The data show that the children with depression did not benefit from taking the drug, the CSM said. It added that there was a two to three times higher incidence of "potentially suicidal behaviour" among those treated with the drug compared with those receiving a placebo, although total numbers were too small to be statistically significant.

The CSM did not rule on using paroxetine to treat social anxiety or obsessive-compulsive disorder, for which there was evidence of efficacy.

On 13 June, Senator Charles Schumer (Democrat, New York) called on the US Food and Drug Administration (FDA), to investigate the effects of paroxetine on children and teenagers. The FDA was expected to announce an investigation this week.

Over the past few years, GSK has been under increasing pressure from patients and their supporters, who say that paroxetine increases suicidal tendencies in all age groups. The company says that there is no evidence for this.

David Nutt, a psychopharmacologist at the University of Bristol, UK, says that the CSM ruling has little bearing on the debate about paroxetine's possible side effects in depressed adults. The CSM was concerned that the drug was ineffective in depressed children, so there was no reason to take any risk, he says. ■



The climate-change debate in Russia is being charged by natural disasters such as this glacier collapse.

Researchers rattled as Kyoto Protocol hangs in the balance

Quirin Schiermeier, Munich

European climate researchers are expressing unease about the objectives of a conference to be held in Moscow this autumn. They fear that sceptics in Russia want to use the meeting to help block ratification of the Kyoto Protocol on Climate Change.

Several scientists, who didn't want to be identified, told *Nature* that they had considered boycotting the forthcoming World Climate Change Conference (WCCC), after attending a programme committee meeting in Moscow last month.

But the researchers said that they had decided to press on with the conference, in the belief that a boycott would offend the Russians and jeopardize climate negotiations inside and outside Russia.

Ratification of the Kyoto Protocol is currently being debated by the Russian administration and, if it happens, would mean that the international treaty would finally come into force. But the debate is very finely balanced, and some researchers fear that climate-change sceptics in Russia want to use the WCCC meeting to block ratification.

Western researchers privately complain about what they regard as the autocratic behaviour of Yuri Izrael, the Russian chairman of the programme committee, at last month's preparatory meeting. It was a "pseudo-democratic exercise, everything was stage-managed," says one German participant.

Izrael is a vice-chairman of the Intergovernmental Panel on Climate Change (IPCC), and a science adviser to the Russian president, Vladimir Putin, who will open the meeting on 29 September. The 73-year-old

former director of the Russian weather service, Izrael is described by his critics as firmly rooted in Soviet scientific traditions. He is also well-known for his reservations about the Kyoto Protocol.

At least one participant has decided not to attend the WCCC. "The objectives of the meeting have not become clear to me at all," says Ulrich Cubasch, a meteorologist at the Free University in Berlin. "I don't want to let myself be used for purposes which I may not want to support."

"The WCCC would be an occasion to raise important issues," says Hervé Le Treut, director of the Laboratory of Dynamical Meteorology in Paris. "But after all the difficulties in discussing things freely at the preparatory meeting, I have very mixed feelings about this conference." He adds that he has not yet decided whether he will accept an invitation to give a talk at the conference.

But Izrael dismisses the criticism. "It will be a purely scientific meeting," he says. "There will be no political decisions or recommendations."

The first scientific speaker on the programme, former IPCC chairman Bert Bolin, has said that he will attend the conference, and most of the 37 invited Western keynote speakers are following suit.

"A boycott of Western scientists would only help those who might be interested in steering the climate debate in Russia in a direction that they see fit," says Hartmut Grassl, an atmosphere researcher at the Max Planck Institute for Meteorology in Hamburg. ■

♦ www.wccc2003.org

Ecologists seek to turn tide on Colorado River

Rex Dalton, El Golfo de Santa Clara

Scientists and environmental groups are battling on several fronts to try to revive the Colorado River delta, a once-vibrant ecosystem in the Mexican desert.

In US courtrooms, in diplomatic meetings and even on the waterways of the delta itself, advocates are trying to secure the primary component that can enhance the region: the flow of fresh water. With almost all of the Colorado's flow extracted for irrigation and water supply in the United States, the river is reduced to a tiny stream by the time it arrives at the Gulf of California in Mexico.

On 4 June, the US District Court in Washington rejected a lawsuit filed by seven environmental groups to force US authorities to let more water flow into Mexico from the Colorado's vast watershed, under the US Endangered Species Act. They claimed that the reduced flow is endangering several species of fish and shellfish.

But ecologists are optimistic that international action can restore the delta's ecosystem. On the same day as the court case, the International Boundary and Water Commission in El Paso, Texas — a joint agency set up by Mexico and the United States — held the first meeting of a binational forum founded to recommend plans to save the delta.

The region will also be on the agenda when senior environmental officials from the United States, Mexico and Canada meet in Washington next week under the auspices of the Commission for Environmental Cooperation — an offshoot of the North American Free Trade Agreement (NAFTA) — to review a list of priority conservation areas.

Advocates want the United States to release about 37 million cubic metres of water per year into the river, with an extra pulse every four or five years if normal floods don't occur naturally.

Meanwhile, researchers from both Mexico and the United States are at work in the area, and are finding that it badly needs more fresh water. Their work is often perilous, because fisherman suspect that the research might damage their livelihoods and drugs smugglers are active in the area.

In March, geologist Karl Flessa of the University of Arizona at Tucson found that the delta's salinity level is nearly twice as high as normal, which is believed to be endangering a clam species, *Mulinia coloradoensis*.

Lorenzo Rojas Bracho, a coastal oceanographer from Ensenada, Mexico, is conducting acoustic surveys to count vaquita (*Phocoena sinus*), a type of porpoise that is under threat from heavy and sometimes illegal fishing in the gulf. "I don't see how we can save the delta without a real commitment from both nations," says Rojas Bracho.

The drive to save the delta hasn't enjoyed



Drying up: Karl Flessa thinks that the decreased flow of the Colorado River is endangering shellfish.

the profile of the decades-long struggle to conserve the Florida Everglades, for example. But it is becoming urgent as economic and ecological stresses threaten to eliminate the last vestiges of a lush riverbank system that existed 75 years ago, before US dams harnessed most of the river's natural flow.

"This is a huge issue," says Lance Morgan, chief scientist at the Marine Conservation Biology Institute in Redmond, Washington.

"We will witness major extinctions in our lifetime if a strong binational agreement isn't reached." Species at risk include clams, shrimp and fish.

Some ecologists are encouraged by a US Bureau of Reclamation plan, whereby the federal government could purchase water from US farmers for release down the Colorado River into Mexico. However, the plan is expected to meet resistance in Congress. ■

Hive beetle causes a buzz in Europe

Philipp Graf, Munich

A new predator could be set to sweep through central Europe's beehives — and angry bee-keepers are asking the European Union to keep it at bay by banning all honeybee imports.

The parasite in question — a small hive beetle — has destroyed thousands of honeybee colonies in the United States over the past five years. Now bee-keepers fear it could find its way to Europe, compounding problems caused by the *Varroa destructor* mite, which wiped out one-third of Europe's bee colonies last winter.

Larvae of the beetle, *Aethina tumida*, which is thought to originate in South Africa, tunnel through honeycombs and kill bee larvae. Infested colonies that do not die often leave their hives. The beetle is hard to eradicate, as it is very mobile and can live in wild bee or bumble-bee colonies, which potentially serve as a natural reservoir.

Since its first identification in Florida in 1998, the beetle has spread rapidly. It has also appeared in Egypt and Australia.

European bee-keepers have been importing queen bees and colonies from the United States and Australia to resupply their colonies after the *Varroa* attack. But this "poses a high risk for bringing in the

beetles", says Jeff Pettis, an entomologist at the Beltsville Agricultural Research Center in Maryland.

German bee-keepers campaigning for an import ban claim that four times as many bee colonies as normal could be imported into Germany this year because of the mite pest losses. And France is drafting legislation for a complete ban on bee imports from affected countries.

Britain already restricts bee imports to queens brought from New Zealand and Hawaii. Now campaigners have won the backing of the German government for a Europe-wide ban on bee imports. A European Commission working group will discuss the proposal in Brussels on 24 June. ■



Unwelcome guest: the beetle *Aethina tumida* has destroyed thousands of US honeybee colonies.

R. DALTON

DOCAS, FLORIDA

Engineers' concerns to be heeded as NASA aims for safer shuttle

Washington NASA has tentatively set mid-December as the launch date for the first space-shuttle flight since Columbia exploded in February. In the meantime, officials say they will instigate several new rules designed to minimize the chances of another accident.

The shuttle will only be allowed to launch during the daytime, for example, and will jettison its external fuel tank in orbital daylight. This will allow cameras to monitor the shuttle during its descent.

A more formalized mission management structure will also be instituted to make sure that any concerns raised by engineers are heard. After the Columbia accident, several engineers said they that they had notified the agency about potential safety problems, but that NASA had not acted on their warnings.

Contingency plans are being considered to use the International Space Station as an emergency haven for astronauts and to keep a sister shuttle on standby before each launch. If all goes well, the shuttle could resume flights to the space station by February next year.

Pacific island tips balance in favour of biosafety protocol

London It's all systems go for the Cartagena Protocol on Biosafety, an international agreement set up to ensure that living genetically modified organisms do not adversely affect biological diversity or human health when moved across country borders.

Last week the Pacific island of Palau became the 50th state to ratify the protocol, reaching the threshold required for the treaty to enter into force. It will now become binding on 11 September this year.

The protocol is designed to ensure that countries are provided with the information they need to make informed decisions about whether to accept transgenic imports. It covers any transgenic organism intentionally introduced into the environment, such as seeds or fish, or that is used indirectly in food or feed, or for processing.

Several major nations will not recognize the treaty when it comes into force. The United States and Australia, for example, have not signed the protocol.

Fraud-buster to be university president

San Diego The University of California (UC) last week named Robert Dynes, a physicist and chancellor of the university's San Diego campus, as its next president.

Dynes will take over in October, when Richard Atkinson retires as president. The

Vincent's golden image of a July night

Paris Astronomers say they have pinned down the date and time portrayed in one of Vincent van Gogh's most famous paintings.

The Dutch master painted *Moonrise* during the summer of 1889 while staying in Saint-Rémy-de-Provence in southern France. But the exact moment depicted in the landscape had eluded art historians until now.

It was 13 July 1889 at 21:08, say astronomer Donald Olson of Southwest Texas State University in San Marcos and his colleagues. To solve the mystery, the team travelled to Saint-Rémy-de-Provence in June 2002 and identified landmarks in the painting. These include an overhanging cliff that obscures a wedge of the luminous orange Moon, and a distant double-roofed house.

Olson and colleagues then worked out where van Gogh was standing when he painted the canvas. Using lunar tables and astronomy software, they calculated the time and dates at which a rising full Moon would appear above the horizon at that spot: 16 May and 13 July 1889. Because the wheat in the painting is harvested, they conclude that it must be the July date. Art historians caution, though, that van Gogh may not have tried to recreate the exact positions of the Moon and other objects.



nine-campus UC system — a tenth campus, at Merced, is due to open in 2004 — employs 160,000 faculty and staff, and has nearly 200,000 students.

Born in Ontario, Canada, Dynes spent 22 years conducting research for AT&T Bell Laboratories in Murray Hill, New Jersey, before becoming chancellor of UC San Diego in 1996. Last year, his lab at San Diego helped to expose the misconduct of former Bell Labs physicist Jan Hendrik Schön, who was found to have fabricated data in a string of high-profile papers on molecular electronics and superconductivity (see *Nature* 419, 419–421; 2002).

Molecular medic to lead cancer research giant

London Cancer Research UK, Europe's largest cancer charity, is to get a new head. Alex Markham, currently director of the Molecular Medicine Unit at the University of Leeds, will take over in September.

Markham replaces Paul Nurse, who leaves just a year after after Cancer Research UK was created by the merger of the Imperial Cancer Research Fund and the Cancer Research Campaign. Since then, income has risen from £238 million (\$400 million) to £270 million, some £10 million above its projections. The organization has also started a collaboration with the American Cancer Society and the World Health Organization, aimed at helping developing countries initiate an international treaty on tobacco control.

Markham, a member of the government's Gene Therapy Advisory Committee, also has extensive management experience in the pharmaceutical industry.

Nurse leaves to take up the position of director of Rockefeller University in New York (see *Nature* 421, 461; 2003).

Incredible! Hulk lab is nothing like reality

Washington When *Hulk* roars to life in cinemas this summer, it will in part be thanks to the hard work of nuclear physicists at Lawrence Berkeley National Laboratory in California.

Set in San Francisco, the movie follows the enraged outbursts of fictional Berkeley physicist Bruce Banner, who gets bigger, greener and meaner after being exposed to γ -rays from the lab's Gammasphere device. The Gammasphere found its way into the movie after producers saw the machine's website. Film-makers took several trips to see the machine at first hand, and a portion of the movie was shot at Berkeley.

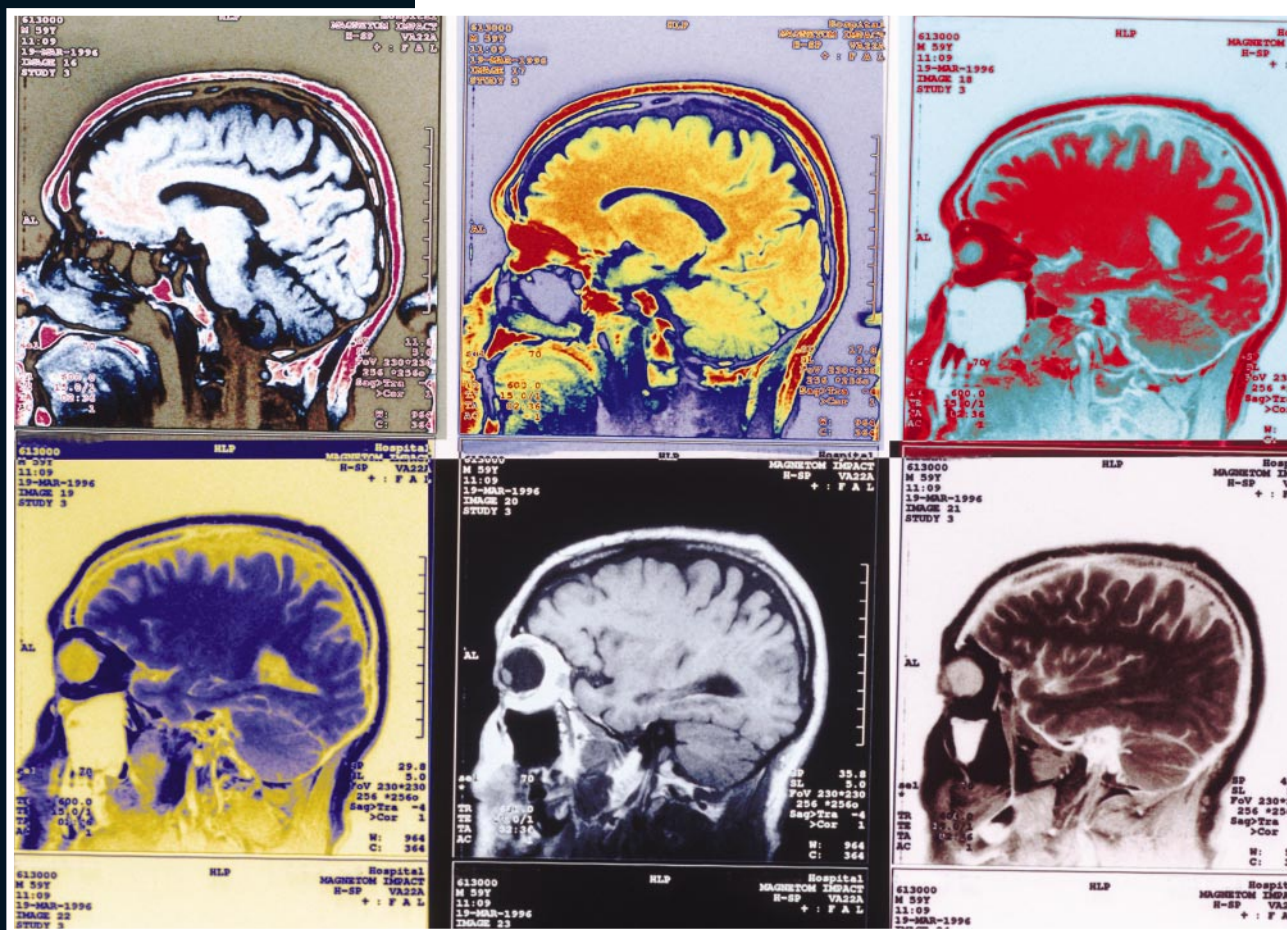
But there are some marked differences between the movie's Gammasphere and the real machine. Most notably, whereas the Hollywood Gammasphere emits γ -rays, the real machine detects them — making it impossible for anyone to be irradiated by the device. "We study deformed nuclei," says Berkeley physicist I. Yang Lee. "We don't deform people."

♦ www-gam.lbl.gov

♦ www.thehulk.com



Green revolution: Hulk discovers a satisfying way of dealing with recalcitrant equipment.



TOM & DEE ANN MCCARTHY/CORBIS

Remote control

Could wiring up soldiers' brains to the fighting machines they control be the future face of warfare? Hannah Hoag investigates the US military's futuristic neuroengineering research programme.

Magnetic resonance imaging lets us monitor the brain, but can we connect directly to neural circuits?

Alan Rudolph sounds like an over-excited sci-fi fan. In his vision of future warfare, the neural circuits of military personnel will be wired into the silicon circuits of the equipment that they control. The brains of fighter pilots, for example, could be connected direct to the controls of their planes. By sending signals to the plane by radio waves, and using cameras on the front of the plane to relay images into the visual areas of the brain, pilots could even control fighter jets from the safety of the ground.

But Rudolph has the clout to do more than fantasize. A project coordinator at the US Defense Advanced Research Projects Agency (DARPA), last year he allocated US\$24 million — almost 10% of DARPA's basic research budget — to a two-year project involving mathematicians, biologists and materials engineers charged with developing technologies to interface brain and machine. Many stumbling blocks, both ethical and technological, will have to be overcome. But agency officials have no doubt about the project's potential. "In the long run, we could have brain-to-brain communication; we could improve the performance of normal healthy individuals," says Rudolph.

DARPA began dabbling in neuroscience in the early 1990s, when the agency grew neurons on silicon circuits and used them to detect chemical nerve agents. In 1998 the programme expanded, developing projects that included research into how honeybees manage to be so manoeuvrable at high speeds. But the agency's neuroscience work didn't attract much public interest until last May, when Sanjiv Talwar, a bioengineer at the State University of New York in Brooklyn, published work on what came to be known as Roborat¹.

Using funding from DARPA, Talwar's team implanted electrodes in the brains of five rats: one in the medial forebrain bundle, which is associated with pleasurable reward feelings linked to actions such as eating or drinking, and two others in the parts of the brain that process signals from the rat's left and right whiskers. Using signals to the reward area as an incentive, the researchers were able to train the rat to move left or right in response to signals to the appropriate whisker area. The rats could then be guided through a maze, along ledges and even be instructed to climb and jump.

The experiment showed a crude degree of communication, using just three electrodes, but it proved that useful information could

be sent directly into an animal's brain. What would be possible, wondered Rudolph, if more connections could be made? Could detailed information, such as images or sound, be beamed into the brain? Could such information be extracted by recording neural signals? And, most important, could it work in humans, so that images, speech and messages could be exchanged between brain and machine, or even between brains?

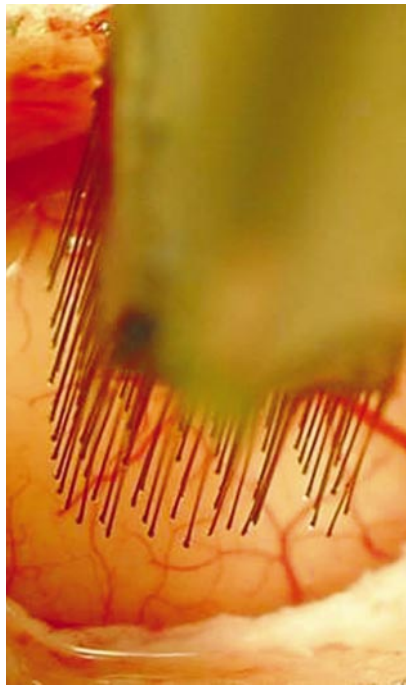
These hopes led to last year's formation of the Brain Machine Interface programme. The scheme is loosely structured into three areas — motor skills, sensory perception and memory. Rudolph identified the neuroscientists, materials engineers and mathematicians who he thought were best suited to the programme. After soliciting proposals from these researchers, a DARPA expert team selected the projects that it believed would bring the agency closer to achieving brain-to-machine and brain-to-brain communication.

Motor motivation

At Duke University in Durham, North Carolina, neuroscientist Miguel Nicolelis and his colleagues are tackling the motor cortex, a brain area that controls movement. The team made a name for themselves, and attracted some DARPA funding, during the mid-1990s, when they taught a rat to control a lever via electrodes implanted in its brain². The rat first learnt to press the lever to earn a drink of water. As the rat pressed, Nicolelis recorded signals from 46 neurons in the motor cortex. He then cut the connection between the lever and the drink. The frustrated rat pressed the bar repeatedly, but received water only when the 46 neurons repeated the same firing pattern that Nicolelis had originally recorded. After a few hours, the rat learned to earn water through thought alone.

When Nicolelis moved his experiments into a tiny owl monkey named Belle at the end of the decade, the implant had grown in complexity and the activity of around 100 neurons could be monitored. As Belle used a joystick to follow a cursor across a screen, a computer compared the movement of the joystick with the signals from the implanted electrodes. After learning how different signals related to commands for speed, direction and force of movement, the computer was, unknown to Belle, able to use her brain activity to drive a robot arm in another room³.

Since then, Nicolelis has started using macaque monkeys, because their brain morphology more closely resembles that of humans. This time, the aim is for the monkey to learn to move a robot arm using signals sent directly from its motor cortex. Seven hundred microelectrodes are positioned in ten different regions of the motor cortex, recording signals from close to 300 neurons.



Mind reader: electrodes are inserted into the cortex of a macaque to monitor neural activity.

At the same time, tiny sensors monitor signals from the monkey's arm muscles, which compare the monkey's intended trajectory with that of the robot arm. "There are a lot of signals being monitored together," says Nicolelis. "On a loudspeaker, it is like listening to a symphony."

The work is currently undergoing peer review for publication, so Nicolelis cannot reveal many details. But, he says, he has been able to send back signals from the robot arm to the monkey's brain. Rather than learning to move the arm by watching it, the monkey can now direct an unseen robot arm using feedback transmitted direct to its brain. Suddenly Rudolph's dream doesn't seem so unrealistic.

While Nicolelis is extracting movement information, other DARPA-funded neuroscientists are working out how to transmit sounds and images into brains. Tomaso Poggio and James DiCarlo at the Massachusetts Institute of Technology, together with Christof Koch at the California Institute of Technology, are attempting to unravel the brain functions that underlie visual object recognition. DiCarlo is targeting a part of the brain called the inferotemporal cortex, thought to be critical for object

recognition, and monitoring the activity of neurons in the brains of monkeys performing visual tasks.

Auditory brain functions are being tackled by Jon Kaas at Vanderbilt University in Nashville, Tennessee, who is trying to decipher the codes that auditory areas use to represent sounds. Kaas is currently recording the responses of different populations of neurons in the auditory cortex of macaques. He says that he can reliably record from about 30 neurons, but that the potential to record from hundreds of neurons exists. "We haven't yet been able to record in large numbers or in the locations we want," he says.

Eventually, both groups hope to have developed enough understanding of how images and sounds are represented in the areas they are studying to be able to send signals back into those areas, creating the feeling that the animals are perceiving real stimuli. "Ultimately the goal would be to electrically stimulate the neurons and see if the animals recognize the sound," says Kaas.

Finally, neuroscientists Sam Deadwyler of Wake Forest University and Ted Berger at the University of Southern California have teamed up to tackle the hippocampus, an area of the brain involved in the storage of memories. They aim to test whether silicon can replace parts of our brain. Their focus is a three-stage chain of regions within the hippocampus that signals are passed through during memory storage. The researchers want to see if they can build a microchip that can take signals from the first region and relay them to the final stage, bypassing the middle part of the chain.

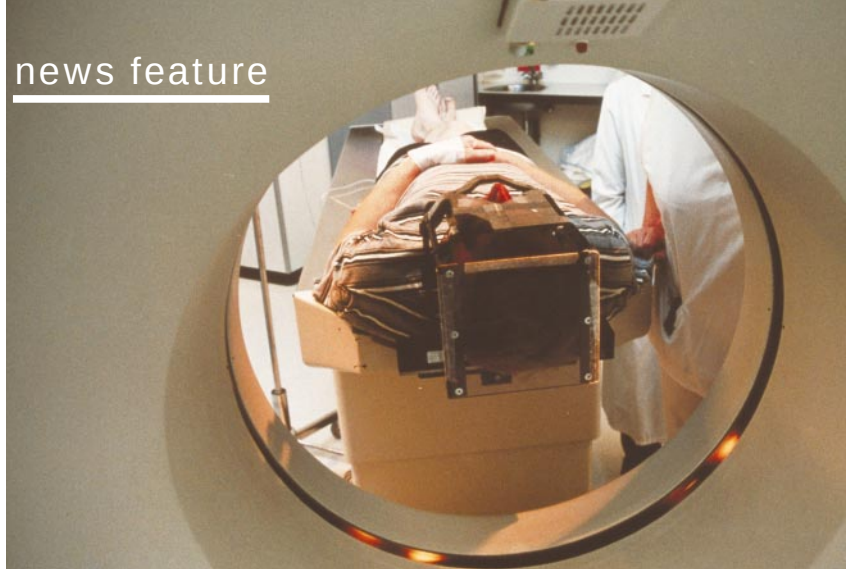
Microprocessor memory

Berger's team has started by developing a picture of what the hippocampus does. By stimulating the neurons in slices of rat hippocampus and monitoring the output that the stimulation produces, they hope to build a mathematical model that mimics processing in the hippocampus. "We'll have a model that will predict how the hippocampus will respond to any input," says Berger. From there, they hope to develop a microchip that recreates that same processing.

Deadwyler has complemented Berger's research by recording the activity of individual neurons in the hippocampus of rats. This work should help the pair decide where best to insert the microchip when it is ready in around two years. One of the hurdles is being able to build the electronics on a micro-scale so that the chip can fit beneath the skull.

Despite the military funding, all the projects have potential medical benefits. DiCarlo's work, for example, will improve our knowledge of the little-understood inferotemporal cortex. Lesions in this region cause agnosia — a rare condition that prevents people from recognizing or identifying objects. He is targeting a part of the brain

The military has always been visionary when funding neuroscience.



Head count: PET scans can monitor brain activity, but not signals from individual neurons.

that has not been tested before. "It's not a new technique: it just hasn't been used in this area," he says. We know very little about the organization of the neurons; that is what makes it a high-risk project."

A synthetic hippocampus, on the other hand, could help those suffering from Alzheimer's disease, stroke and epilepsy, says Berger. "The malfunctioning hippocampus is difficult to care for," he points out.

Scanning signals

But Berger is well aware that DARPA's goals go beyond substituting a chip for neurons. Although the researchers involved seek therapeutic devices, each device could, in theory, be used to enhance normal human function. In the long run, Rudolph hopes that these different strands of work will give birth to technologies that allow images to be relayed direct to the brains of military personnel. When they decide how to act on the information they receive, electrodes in their brains will decipher their decisions and relay them to the minds of their colleagues.

Pilots, for example, could receive sensory input and manoeuvre their plane through thought alone. Berger envisages a scene reminiscent of *The Matrix*, in which a complex evasive manoeuvre could even be programmed into future versions of his memory implants, allowing the pilots to perform moves they may not actually have learned through traditional training. Berger adds that the system could detect that a pilot was about to make an error and override his actions. "When you are flying an F-18 you can't afford to make a mistake," he says.

But are such goals anything more than the fantasies of military and scientific futurists? In the immediate future, technological problems may block progress. The brain views an implanted electrode as a foreign invader and sends cells to encapsulate the electrodes in tissue, preventing signal transmission. By using flexible, teflon-coated electrodes, Nicolelis has managed to keep electrodes implanted in the brains of the monkeys for

two years and maintain the integrity of the recorded signal. Even so, he says that more work needs to be done on the biocompatibility of different materials.

In the longer term, developing a non-invasive method of recording brain signals could be a bigger stumbling block. "We don't have any plan or interest in taking the devices we have now and implanting them into healthy people," says Rudolph. Existing non-invasive systems are not up to the job. Medical scanning procedures such as magnetic resonance imaging (MRI) and positron emission tomography (PET) have provided enormous insight into brain functions. But MRI measures blood flow and PET traces the movement of molecules tagged with a radioactive isotope — neither has the resolution to monitor the activity of individual brain cells. "There is nothing that directly measures the electrical signals that flow back and forth between neurons," says Rudolph.



Thought control: Miguel Nicolelis with Belle and the robotic arm that can mimic her movements.

The solution may come from future projects headed by materials engineers. "DARPA is good at mixing up communities of disciplines," says Rudolph, who adds that if a microchip and electrodes could be implanted into the brain without harm or pain, there might not be a need for a non-invasive device. "But these are things that we can't envisage right now," he says.

While DARPA wrestles with these long-term problems, the neuroscientists involved are simply pleased to be able to carry out such audacious research. Most say that if it weren't for DARPA, they would not be doing the work they are now. "The military has always been quite visionary when funding neuroscience, whereas the National Institutes of Health (NIH) has been conservative," says Koch.

Whose responsibility?

In effect, each agency approaches neuroscience from the opposite perspective. The NIH pursues brain science from the ground up, and funded many of the early projects of these researchers. "At the NIH you do incremental science, taking little bitty steps," says Berger. DARPA, by contrast, prefers riskier projects with higher payoffs. "We hope that by building the interfaces we will learn how it can be done," says Rudolph.

But money from DARPA can come with different problems. Unlike NIH projects, which may be funded for five years at a time, researchers working for DARPA could find their funding pulled out from under them less than two years later, leaving employees and graduate students stranded. "You begin to feel insecure," says Kaas.

More important, some feel that using money from the military is ethically problematic. Many of the researchers funded by DARPA feel that the agency's goals are a little fanciful, and none believe that the military is close to achieving its ultimate aim. Nonetheless, several of the researchers that *Nature* spoke to were reluctant to discuss military applications, saying they would rather see their research used benevolently. For some observers, this is an abdication of the responsibility that researchers have for their work.

Martha Farah is a neuroethicist and director of the Center for Cognitive Neuroscience at the University of Pennsylvania, Philadelphia. She argues that a researcher who accepts funding from the Department of Defense, but does not believe in the goals behind the funding programme, is compromising their ethics. A great deal of basic research will be beneficial for some and detrimental to others, especially if the information can be used by the military. "Researchers need to take some responsibility," says Farah. ■

Hannah Hoag is an intern with *Nature* in Washington.

1. Talwar, S. K. et al. *Nature* **417**, 37–38 (2002).
2. Chapin, J. K. et al. *Nature Neurosci.* **2**, 664–670 (1999).
3. Wessberg, J. et al. *Nature* **408**, 361–365 (2000).

The crystal's in the mail...

Do you need to solve the structure of a complex biological molecule quickly, or require some expert help doing it? No problem. Tracy Smith Schmidt samples the delights of 'mail-in' crystallography.

Some say that today's biology graduate students are spoiled. Not so long ago, they had to do all their own routine bench work — from sequencing DNA to making transgenic mice. But core facilities at many universities now provide these services reliably and cost-effectively. Budding researchers, freed from this treadmill, can concentrate on more interesting experimental work.

The next generation could be even more pampered. If they need to solve the crystal structure of a protein, they may not have to master every aspect of X-ray crystallography. Instead, biologists might simply mail crystals of the protein to a specialist centre, where expert staff will perform the X-ray analysis and send back the data electronically. Indeed, some pioneering facilities are already offering such 'mail-in' services.

Solving the structure of a protein, or any other complex biological molecule, is a laborious task that used to take years. First, the protein has to be purified from cell culture, before the painstaking process of finding conditions to get it to crystallize from solution. The crystals are then irradiated with X-rays, to see how the beam is scattered by the crystalline array of atoms. From these data, a three-dimensional map is generated that shows the location of each atom within the molecule.

Booking time

Although the basic steps haven't changed over the years, the time it takes to solve a structure has, dropping to only a few months in many cases. Many technical advances have helped turn this once arcane pursuit into a staple technique of modern biology. In particular, synchrotron facilities — machines in which electrons moving at close to the speed of light are steered by magnetic fields around a huge, ring-shaped accelerator to generate high-intensity X-ray beams — have made many more experiments possible. For example, high-quality data can be collected from tiny crystals that were once deemed unusable.

But booking time to collect data at a synchrotron beamline is still a limiting step — and for crystallography novices, an extremely daunting one. Fortunately, there's now an alternative. Since 2000, the National Synchrotron Light Source at the Brookhaven National Laboratory in Upton, New York, has offered a



At your service: Howard Robinson examines a sample sent to Brookhaven for structural analysis.

mail-in data-collection service: you supply the crystal and they deliver the data.

Run by Howard Robinson, the Brookhaven programme contributes to more than 50 projects each year, offering its services free to academic scientists. Researchers whose proposals are accepted send in their crystals. Robinson's experienced crew collects the X-ray diffraction data, often making recommendations about how best to perform the experiment. Finally, they forward minimally processed data to the user. Unless everyone agrees to collaborate further, Robinson's group does not calculate the final set of atomic coordinates that allows researchers to display all the twists and turns in the protein.

About half of Brookhaven's applicants have been molecular biologists with little previous crystallography experience. The rest have been established structural biologists who are in a hurry to get something done, and who are more than happy for Robinson's team to help them out.

The high demand for synchrotron time

means that crystallographers typically have to apply for a slot on an X-ray beamline months in advance. When they travel to the synchrotron, they must often run their experiments in marathon sessions of 12 to 24 hours. And if they fail to obtain all the data they need, it's back to the end of a long queue. In contrast, Robinson's mail-in service usually delivers results within a few weeks of submitting a proposal.

For non-crystallographers, access to expertise, rather than speed of service, is a major draw. Ziwei Huang, a biochemist at the University of Illinois at Urbana-Champaign, studies chemokines, proteins involved in coordinating the immune response. He had long been interested in pursuing structural work but was unsure how to jump into the field. Over the past couple of years, his team has made the leap, thanks to the Brookhaven mail-in service and to a facility on Huang's home campus that helped his graduate student prepare crystals of the chemokines they wanted to analyse.



Elspeth Gordon (right) at the European Synchrotron Radiation Facility (above): automation will soon replace the human touch.



"As I did not have much prior experience in this area, without these services it would have taken us much longer to go in this direction," says Huang. Having been introduced to the world of crystallography, his group now has the confidence to pursue independent structural studies.

But taking advantage of expert knowledge in this way does introduce complications — not least the thorny issue of how the experts should be credited for their work. Robinson says that he expects members of his team to be offered co-authorship of papers if they make a "significant contribution" to a project. But there are no hard-and-fast rules. "It's a grey area," says Robinson. "We try to decide the needs of the project, and users have to decide what is a major contribution."

Express service

The Brookhaven group is not alone in offering mail-in data collection. At the European Synchrotron Radiation Facility (ESRF) in Grenoble, France, Elspeth Gordon and Stéphanie Monaco run an operation called MXpress, which provides a mail-in service for a fee to anyone willing to pay.

For Vincent Mikol, head of structural biology with Aventis Pharma in Paris, MXpress has freed his staff from the travel and unsociable hours associated with working at a synchrotron facility. "We really wanted to avoid spending time, specifically so many nights, at the synchrotron," he says. So a year ago, Aventis entered into an agreement with Gordon and Monaco that has so far provided data on more than 200 biological samples.

Mikol is pleased with the results, but acknowledges that staff unused to the idea of

trusting a third party to handle their precious samples had to overcome some mental roadblocks. "At first people were reluctant to send new crystals, but after a while, they saw that the ESRF group was responsive and capable," says Mikol. "We still travel to synchrotrons a bit, but the goal is to have this fade out completely in the next two years."

Trust is an issue for other researchers as well. "If the crystallization were straightforward and reproducible, we would definitely be keen to try a service such as Brookhaven's," says Brenda Schulman, a structural biologist at St Jude Children's Research Hospital in Memphis, Tennessee. "But using a mail-in service for projects where every crystal is extremely precious could put us both in an awkward situation if something were to happen to the samples or if the data collected were not optimal."

These concerns may lessen within the next few years, as data collection becomes fully automated. Machines, not technicians, will pick up the crystals, mount them on the apparatus, centre them in the X-ray beam, and collect the data. Bar-coding samples should further reduce the scope for human error.

In addition, remote-access software is being developed to allow scientists to observe and manipulate operations at a synchrotron

We wanted to avoid spending so many nights at the synchrotron.

Vincent Mikol

from their own labs. Users will be able to monitor their data as they appear and make judgement calls about how to proceed — for example, to request a move to another crystal if the one in the beam is of poor quality. Once this capability is in place, scientists will be able to stay intimately involved in the execution of their own experiments — they will just be physically removed from the site of data collection. The major issue to overcome is security: how to create firewalls to prevent hackers from gaining access to a user's results and to the beamline equipment.

Keeping in touch

But even when this technology is available, some researchers may still prefer the hands-off approach of having the synchrotron staff run their experiments. Some scientists raise the question of whether new recruits will suffer in the drive for greater efficiency and miss out on learning the intricacies of crystallography. "We run the risk of not training graduate students at the beamline," says Stephan Ginell, who coordinates the Structural Biology Center User Program at the Advanced Photon Source, a synchrotron at the Argonne National Laboratory in Illinois.

Most structural biologists are confident that a balance can be struck. Ashley Deacon of the Stanford Synchrotron Radiation Laboratory in California, for one, believes that many researchers will want to stay hands-on — which should ensure that graduate students get the training they need. "For the most part, researchers still want to be highly involved," he says. "A lot of human interpretation still goes into it."

Even so, it may eventually become standard for biologists to contract out crystallographic procedures. Already there are companies, such as Rigaku/MSC of The Woodlands, Texas, offering the full gamut of services, from crystallization to structure solving.

As more outsourcing opportunities emerge, the issue of training may become moot. "I used to spend a great deal of time as a graduate student sequencing DNA, but now this is done as an automated service," observes Schulman. "Mastering particular protocols may be less important than learning the process of how to do good science." Still, she notes that researchers who rely solely on others to perform their structural analysis may risk overlooking errors and artefacts in the data.

As this debate continues, the signs are that demand for mail-in crystallography is growing. Those synchrotrons that haven't yet offered the service are getting requests to do so. "I get calls from people demanding it for general users," says Ginell.

Tracy Smith Schmidt is a freelance writer in New York and was formerly editor of *Nature Structural Biology*.

♦ www.px.nsls.bnl.gov

♦ www.esrf.fr/Industry/MX/MXpress

How e-mail raises the spectre of a digital Dark Age

Records can be wiped out by a click of the mouse or rendered unreadable by upgrades.

Sir — Throughout the ages, historical documents have been carved in stone, etched in clay, written on parchment and printed on paper. Examination of traditional documents requires merely the ability to read and understand the language or code in which they were written. Saving them requires but a thoughtfully selected repository, with a security device if needed. But the term “read” has a completely different connotation when referring to a floppy disk or a computer hard drive. So, whereas the Rosetta Stone only required interpretation of the written key, today’s equivalents require all the correct operating systems and software applications.

Faced with these complexities, it is all too tempting to dump records in the recycling bin. It is nothing short of terrifying to contemplate the probable magnitude of lost records as yesterday’s electronic storage devices become incompatible with today’s software applications.

These problems are most alarming when considered in the context of electronic mail. Worldwide, the number of corporate mailboxes is projected to grow from 131 million in 2001 to 225 million in 2005; daily e-mail traffic is anticipated to grow from 9.7 billion pieces in 2000 to more than 35 billion in 2005.

Little of this staggering volume is translated to hard copy. Even formal documents are frequently discarded by simply erasing them to avoid the tedium and expense of printing and archiving, on the erroneous assumption that previous iterations are of no value. The threat to the existence of hard-copy records of e-mail will become more acute as younger generations become less culturally and educationally attuned to their importance for the preservation of historical records.

A perfect electronic world should not require hard copies of e-mail (or any other electronic document). E-technology should have addressed this problem as it evolved. But as individuals, institutions and corporations have continually ‘upgraded’, it has become more problematic to preserve e-documents because formats, servers and institutional IT policies change. The US National Archives has recently acknowledged the lack of robust e-mail archiving solutions, but has determined that long-term storage of electronic communications must be independent of these variables (see www.archives.gov/records_management).

Institutional policies also negatively affect long-term retention of e-mail. With limits of hard-drive space on mail servers,

institutional users are routinely faced with the requirement to cleanse mailboxes. The pen may be mightier than the sword, but a single mouse-click can destroy products of inestimable value.

The ease and simplicity of e-mail has altered the fundamental psychology of communication — and of confidentiality. Most of us are oblivious to the reality that e-communication violates the confidentiality intrinsic to conversations: e-mail records (even deleted ones) can be subpoenaed by courts of law. Other people’s computers can easily be used to send or receive e-mail, or to access the Internet. The potential for violations of confidentiality and for impersonation is all too obvious, and examples of ‘e-fraud’ abound.

Even in honest use, the immediacy of e-mail breeds poor communication and can rapidly escalate conflict. The facile nature of the medium encourages a lack of

sensitivity, and the false assumption that such communication is private and confidential can lead to situations such as that of the recovery of e-mails questioning the safety of the US space shuttle’s re-entry written days before the recent disaster.

Electronic communication is surely here to stay. Its simplicity and enormous speed will ultimately send other modes of communication close to extinction. But IT solutions to the many problems it brings are long overdue. Most serious is the potential for the loss of records that may have immense historical value. For a start, academic (and other) institutions should require hard copies of certain categories of electronically generated information to be kept — with appropriate authentication.

Errol C. Friedberg, Herbert K. Hagler, Kevin J. Land

Department of Pathology, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA

Britannica: source of a passion for physics

Sir — In his obituary of US physicist Julian Schwinger (*Nature* **370**, 600; 1994), David S. Saxon wrote: “He once remarked that he had been reading straight through the *Encyclopaedia Britannica* and when he came to the letter P and to physics, that was it.”

In another memorial article, Paul C. Martin and Sheldon L. Glashow wrote “Armed with the 11th edition of the *Encyclopaedia Britannica* and the books and journals in nearby libraries, Schwinger set himself apart from the establishment of teachers, textbooks and assignments.” (*Physics Today* **48** (10), 40–46; 1995).

These descriptions raised my interest in the contents of the famous 11th edition of the *Encyclopaedia Britannica* (1910), which also influenced the physicist Freeman Dyson profoundly (see S. S. Schweber *QED and the Men Who Made It: Dyson, Feynman, Schwinger, and Tomonaga* 479; Princeton Univ. Press, 1994).

After examining this edition, I found to my surprise that ‘physics’ is not an entry word. This is particularly strange as this edition was published a decade and a half after the great discoveries that marked the beginning of modern physics. In contrast, both ‘chemistry’ and ‘physiology’ are entry words, with many depictions of these two sciences. Was Schwinger’s account of his early commitment to physics perhaps either a joke or a fable?

Of course, many entry words regarding

the science of physics were included and embedded in different volumes of the *Encyclopaedia Britannica*, so one could well believe that young Schwinger learnt physics from the entire set.

Min-Liang Wong

Department of Veterinary Medicine, College of Veterinary Medicine, National Chung-Hsing University, Taichung 402, Taiwan

Britannica: which edition inspired Schwinger?

Sir — Regarding Julian Schwinger’s story of finding his subject by reading through the *Encyclopaedia Britannica* and stopping at ‘Physics’: an entry entitled ‘Physics’ appeared in every edition, beginning with the first in 1768, until the 8th (1852–60). In the 9th (1875–89) it was replaced with ‘Physical Sciences’ (by James Clerk Maxwell), from a paper “Remarks on the classification of the physical sciences”. ‘Physics’ as an entry title did not appear in the 11th or 12th, but it did appear in the 13th (1926), and in the modern sense, whereas previously it was an alternative name for ‘natural philosophy’ or ‘the science of nature’. The author was Robert Andrews Millikan. The 12th edition was actually the 11th with some supplementary volumes, as was the 13th.

Anita Wolff

Senior Editor, Encyclopaedia Britannica, 310 South Michigan Avenue, Chicago, Illinois 60604, USA

A walk to the gallows

Could scientific advances be hastening the end of the world?

Our Final Hour: A Scientist's Warning: How Terror, Error, and Environmental Disaster Threatens Humankind's Future in this Century – On Earth and Beyond/ Our Final Century: Will the Human Race Survive the Twenty-First Century?

by Martin Rees

Basic Books: 2003. 256 pp. \$25/

William Heinemann: 2003. £17.99

Don Brownlee

I must admit I was startled when I first heard the title of Martin Rees' new book, *Our Final Hour*. Is it really possible that our end could be so near? Could our species become extinct less than three centuries after the publication of Darwin's *Origin of the Species*? We humans have only just begun to understand what we are and how we fit into the cosmos. How could the curtain fall on us so soon? My initial shock and knee-jerk scepticism was eased, at least a bit, when I found that the UK version of this book is titled *Our Final Century*. As is the case with most apocalyptic assessments, the prediction may involve a few orders of magnitude of uncertainty.

Gloom and doom is an intriguing and popular genre but it can also be disturbing. Together with palaeontologist Peter Ward, I recently wrote *The Life and Death of Planet Earth* (reviewed in *Nature* **422**, 663; 2003). In writing the book, and in numerous talks and interviews given after its publication, we had to deal sensitively with people's fears of an unpleasant, albeit distant, future. As an astronomer and a palaeontologist we feel comfortable with the premise that the Earth will lose its plants and animals in a billion years or less, its oceans and all life within several billion years, and ultimately become assimilated into the red giant Sun in about 7 billion years. We are at peace with our cosmic fate, smugly encapsulated in the present and well isolated from the future. But many readers are genuinely disturbed that there are bad things ahead, even if they are millions or billions of years away.

Now Martin Rees, Britain's astronomer royal and an author entitled to be taken very seriously on such a weighty topic, suggests that our end is essentially here. He wagers that humans have a 50% probability of becoming extinct by the end of this century. Ward and I had conjectured that humans, being wily, adaptive creatures perhaps immune to future natural selection, might be among the last of Earth's animals to reach extinction. Rees suggests just the opposite: that nearly all of Earth's large mammals may outlive us for the



Virtual reality? Computing advances leave us open to the risk of cyber-doom, as in *Lawnmower Man 2*.

simple reason that they do not have science and technology, and so lack the tools to do themselves in. We cry "save the whales", but we may be the ones who need saving.

There are some profoundly unsettling aspects to this book. It is bad enough that the end might be near, but it is even worse that the cause of our demise could be science. As a scientist, I have always considered science to be the most successful human endeavour. The bootstrap aspect of science and the role of nature as the ultimate arbiter of truth have led us far. Starting with sticks, fire, flint and pottery, our ambitions to comprehend and utilize nature have led to indoor plumbing, spacecraft, the Internet, a sophisticated view of the cosmos, and even a substantial glimpse into the intricate workings of biology.

If Rees is right, the continual advance of science has been a walk to the gallows. Our hard-earned knowledge of natural processes has led us to the threshold of destruction. Within the next century, the ability to build designer viruses, nano-robots or even some type of computer-instigated cyber-doom will have advanced to the point that perhaps a single deranged individual will have the ability to kill us all. There will always be lunatics, so if the power is there, we are doomed. This is the fundamental thesis proposed by Rees.

The book is a shopping list of how the end might come. It includes exotic concepts of how our fiddling might end the world or even muck up the entire Universe, but the

focus is on our self-extinction. Having survived predictions that the 'electronic brain' and the robot would have taken over long ago, my personal guess is that cyber-threats and killer nanobots will never be able to take total control, but who knows?

The threat of specifically designed killer viruses or diseases seems more credible, and new biological threats are sure to cause havoc in the future. But can even the most deadly designer bugs cause total human extinction? The nastiest viruses are usually not very successful because they kill their host before they can be transmitted. Life is actually pretty robust, the product of the tough taskmaster of evolution. Viruses have been attacking bacteria in the oceans for billions of years and, even though they outnumber their hosts by orders of magnitude, the viral attackers can never completely win.

In my mind, humans are nearly extinction-proof unless they are exposed to global changes to which they cannot adapt. If a nanobot, or a microbial or nuclear attack, was only 99.9999% efficient, the remaining isolated patches of humanity could live off the largesse of the modern world for quite a while. The first serious global attacks on humans are likely to be much less than 100% efficient. Even a minor kill-off would provide quite an incentive for the survivors to prevent future attacks.

A quite remarkable aspect of this book is that the author, a famous scientist, clearly assigns the blame for the anticipated

apocalypse on science. The core of the threat is not only science but also knowledge. The book rings the alarm but provides no answers because there are none. Scientists are motivated by curiosity and a yearning to understand nature, but the real engines of science are business and war. Without global cultural revolution, business, war, science and education will surely continue. Can we manage the dark side of science and prove Rees wrong? Only time will tell.

I highly recommend this provocative and educational book. It is written for broad appeal and even includes a crowd-pleasing array of references to popular culture. If we do survive to see the twenty-second century, I hope that Martin Rees' cautionary and alarmist book will be required high-school reading, a twenty-first-century analogue of Rachel Carson's *Silent Spring*. I also recommend *Future Evolution* by Ward and Alexis Rockman (reviewed in *Nature* **416**, 125–126; 2002), a discussion of the fate of humans that favours their survival but warns of dire effects on the rest of the planet's inventory of plants and animals. ■

Don Brownlee is in the Department of Astronomy, University of Washington, Seattle, Washington 98195, USA.

Four legs bad, two legs good

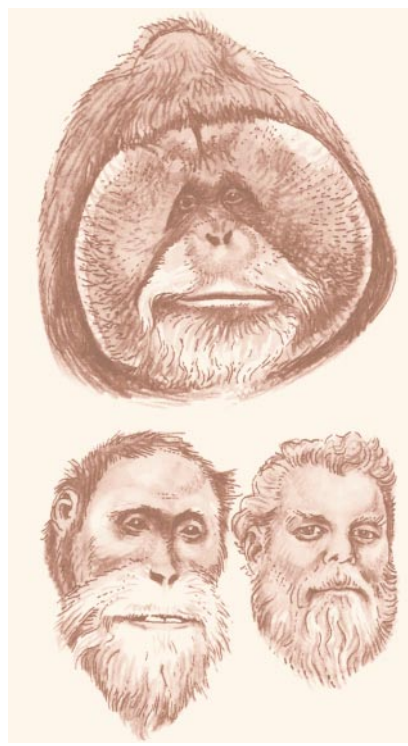
Lowly Origin: Where, When, and Why Our Ancestors First Stood Up by Jonathan Kingdon
Princeton University Press: 2003. 408 pp.
\$35, £24.95

Lorenzo Rook

Jonathan Kingdon is not only a zoologist with a deep knowledge of African mammals, but also a fine narrator and a talented artist, so any new book by him is particularly welcome. In *Lowly Origin* he offers the reader a finely illustrated book on human evolution, focusing particularly on the transition from four-legged apes to two-legged hominins.

The origin of our ability to walk on two legs is one of the most debated subjects in anthropology. Here Kingdon attempts to explain to a wide audience, and to shed new light on, the origin and consequences of this turning point in human evolution. The book's strength is its broad perspective, bringing ecological, palaeoenvironmental and palaeogeographical evidence into an analysis of our "lowly origin" (the title is taken from the last two words of Darwin's *The Descent of Man*).

Over ten chapters, Kingdon provides views on the evolutionary trajectories of forelimbs since Palaeozoic tetrapods, ponders the



Face facts: Jonathan Kingdon highlights the similar expressions of humans and orang-utans.

evolutionary patterns of African hominins as driven by physical limits between geographic regions (by 'basin evolution'), and discusses the role of brain expansion in driving the modern diaspora. He repeatedly refers to "self-portraits", both as an abstract mental exercise comparing humans and other primates or animals, and by including illustrations of himself in comparison with other primates and humans — the example shown above is just one of many intriguing juxtapositions. As a "repentant vandal", Kingdon summarizes in the final chapter our status as the last survivors of the many bipedal African mammals that persisted until recent times, and offers insight into the future of our species.

There are a few weak points, however. In some places the ongoing debate on some crucial issues is missing and a definite position is taken instead. For example, the recently described *Sahelanthropus* fossil from Chad is mentioned as a gorilla-like ape but is totally overlooked in the chapters that discuss the earliest hominin.

Another case is the origin of the African ape–human clade (the taxonomic group including *Homo* and his nearest living relatives, *Pan* and *Gorilla*). Some recent papers have argued that the European Late Miocene apes were the ancestors of the African ape–human clade, suggesting that hominids left Africa for Europe about 20 million years ago and returned some 10 million years later. Kingdon accepts this 'out of Africa and back' hypothesis — the Eurasian origin of his

"squatting ground ape" is a recurrent motif throughout the first part of the book. But in doing so, he neglects the wide palaeoecological perspective that is otherwise the main strength of this book.

The European hominoid with the widest temporal and geographical range is *Dryopithecus*, a large-bodied arboreal ape adapted for a diet of soft fruits. A slightly more recent European ape is *Ouranopithecus* (also known as *Graecopithecus*), a hominoid about the size of a female gorilla that is adapted to eating hard fruits. *Dryopithecus* is associated with swampy subtropical environments (from Spain to Eastern Europe), whereas *Ouranopithecus* (from Greece) and close relatives from Turkey lived in rather younger, more open woodland.

Does the available palaeontological and palaeoecological information support the dispersal of *Dryopithecus* or *Ouranopithecus*-like forms from Europe into Africa? The 'out of Africa and back' hypothesis is essentially based on 'strong' cladistic analyses (a procedure for organizing the evolutionary relationships among taxa on the basis of shared anatomical characters) that represent *Dryopithecus* and *Ouranopithecus* as being sister taxa of the African ape–human clade. This hypothesis is favoured by some authors because there was a 'hominoid vacuum' in Africa between 12.5 million and 6 million years ago. But the large-mammal fossil record in Africa is relatively poor for this time interval, so the absence of hominoid fossils cannot be taken as evidence that hominoids were not living in this area during this period.

Dryopithecus was dependent on equable, subtropical temperate forests, below-branch locomotion and a soft-fruit diet, so it is not likely to have moved across the more open woodland of the southwestern Mediterranean during the Late Miocene. *Ouranopithecus* is also an unlikely ancestor of the African ape–human clade, because its teeth and jaws had thick dental enamel and were adapted for a diet of hard fruits and seeds. As such they were more comparable to the teeth of later mid-Pliocene to early Pleistocene hominids (between 3 million and 1 million years ago) than to the thin-enamelled teeth of the early bipedal hominin of Ethiopia or the earliest Late Miocene hominin from Kenya or Chad.

Despite its lack of discussion of these issues, *Lowly Origin* should interest a wide audience. It will serve as supplementary reading for researchers and graduates, but should also be accessible to lay readers. Kingdon is an authority on Africa's zoology, and has merged his deep knowledge of mammal evolution with his love for Africa and art. ■

Lorenzo Rook is in the Department of Earth Sciences and the Natural History Museum, University of Florence, via G. La Pira 4, 50121 Firenze, Italy.

Towards a biology of the mind

Brain-Wise: Studies in Neurophilosophy

by Patricia Smith Churchland

MIT Press; 2002. 438 pp. \$65, £43.50 (hbk); \$25, £16.50 (pbk)

Alva Noë

The main aim of philosophy is to understand our nature and our place in the universe. In *Brain-Wise*, Patricia Smith Churchland provides an introduction to what she calls 'neurophilosophy' — philosophy as it is being transformed by advances in neuroscience.

The book comprises a series of essays written for the non-specialist on philosophical topics such as the self, consciousness, freedom of will, knowledge and the existence of God, and touches on learning, language and causality. Each chapter contains a survey of philosophers' arguments, and a review of relevant work in cognitive neuroscience. For example, in the chapter on the self, Churchland discusses a variety of neurobiological disorders associated with the breakdown of the self, such as parietal lesions, anorexia nervosa and the dementias. The chapter on consciousness contains an admirable critical review of work on the neural correlates of consciousness. Neuroscientists will find these discussions illuminating, and philosophers will find them informative.

The book is intended as an introductory textbook for students, but it is more than this. It is also an extended criticism of philosophy as seen from the viewpoint of neurophilosophy. Churchland seeks to establish that neuroscience is relevant to at least some philosophical problems, a claim that one might have thought didn't need establishing. This is where the book is most fascinating, but also where it is most open to criticism. The discussion turns on the status of neuroscience as science, and also on the relationship of philosophy to science in general.

Are there any philosophers who deny the relevance of brain science to philosophy? Churchland's main target seems to be those who defend the 'computer model' of the mind, which views the mind as software to the brain's hardware. According to this view, to understand the mind we must ignore the details of the way it is implemented in the brain. Churchland is right to say that the hardware–software distinction has no basis in biology. But as a criticism of the computer model, this observation lands wide of the mark.

There is not much doubt that mental function resides in the brain. The difficulty has always been understanding mental function in neural terms. Neural activity itself — at whatever scale of organization — seems to be

Exhibition

The medicine man's museum

Henry Wellcome is best known for his endowment of The Wellcome Trust, an influential UK research charity. But in the latter part of his life the pharmaceutical entrepreneur used his immense wealth to support his passion for collecting. He scoured the globe for artefacts relating to the history of medicine, collecting any type of object from any era.

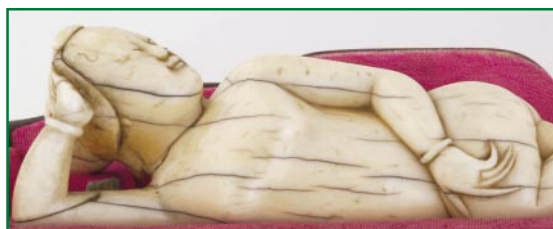
The 700 pieces in the exhibition *Medicine Man: The Forgotten Museum of Henry Wellcome*, which runs from 26 June until 16 November, barely scratch the surface of this vast but little-known collection; the remaining million or so objects are scattered in museums around the world.

The two dolls shown here

exemplify cultural differences in the attitude to the female body. The seventeenth-century ivory anatomical model (lower picture) of a pregnant woman with removable organs, probably from Germany, is thought to have been used by obstetricians and midwives to reassure pregnant women. The eighteenth-century 'diagnostic doll' from China (top) was used by women of wealthy families to indicate where their symptoms lay, as they were forbidden a physical examination by male doctors.

Other objects on display include Vincent van Gogh's only etching, *L'homme à la pipe: portrait du Docteur Gachet*, which is a patient's view of a doctor. Gachet, a proponent of electrical treatment for melancholia, was van Gogh's last doctor.

Alison Abbott



Dolls have served medical purposes in China (top) and Germany.

cognitively inscrutable. David Marr captured this idea in his book *Vision* (W. H. Freeman, 1982) with his remark that "trying to understand vision by studying only neurons is like trying to understand bird flight by studying only feathers: it just cannot be done."

To even begin to understand how neural events enable or constitute mental processes, those processes must be modelled in terms that are abstract in relation to neural implementation. This is certainly the dominant research strategy in cognitive neuroscience today, and one that Churchland herself uses. Consider, for example, her argument that brains use models of the effects of body movement (emulators) to enable sensory-motor coordination and offline planning of action. Is this a claim about brains as such, or is it instead about high-level cognitive architecture? It is striking that Churchland's treatment focuses, for the most part, on general engineering considerations about the utility of emulators.

One of Churchland's arguments for emulators is problematic. As a result of eye movements, the retina registers nearly continuous shifting in light patterns. How does the brain distinguish this eye-dependent change from that produced by the movement of perceived objects? Churchland's answer: "The brain undoubtedly uses emulators."

The idea is that emulators are used to keep track of, and so compensate for, eye movements. The trouble is not with the answer, but with the question, which rests on the assumption that movement of the image relative to the retina is the brain's code for movement in the environment. If we give up this assumption, there is no need for a compensatory mechanism, and hence no work for emulators to do. Emulators in the brain (at least as far as this argument shows) are merely reflections of *a priori* assumptions about cognitive organization.

I am not trying to call the emulator hypothesis into question. Nor do I fault Churchland for pressing the importance of developing a distinctively biological theory of mind. She is right about this. We are not there yet, however. The development of new technologies to study the brain — non-invasive imaging techniques, such as functional magnetic resonance imaging, for example — has done little to enable us to overcome the cognitive inscrutability of neural systems. Progress in this area, as in other fields, is likely to depend on collaboration between philosophy and science. Churchland's book is a step in the right direction. ■

Alva Noë is in the Department of Philosophy, University of California, Santa Cruz, California 95064, USA.

Charmed by snakes

Summarize yourself in the form of a title of a paper in Nature.

Working on a mystery and going wherever it leads.

What was your first experiment as a child?

I had snakes as a kid — a king snake, a yellow snake, green snakes, several kinds of garter snake, and a tiny species called DeKay's snake. I experimented a lot with their choice of food items. The beautiful patterns of some of these species had a lot to do with my future aesthetic interests in biology.

Who has been the most important mentor in your career?

Matt Scott (now at Stanford) gave me a huge, golden opportunity in the early days of molecular developmental genetics and an unusual degree of freedom.

Whose graduate student would you most like to have been (historical impossibility notwithstanding)?

Someone from the 'Morgan school': Calvin Bridges or Alfred Sturtevant. Or Theodosius Dobzhansky.

What single scientific paper changed your career path?

Ed Lewis's review on the Bithorax complex (*Nature* **276**, 565–570; 1978).

What book has been most influential in your scientific career?

A tie between *Ontogeny and Phylogeny* by Stephen Jay Gould and *Embryos, Genes, and Evolution* by Rudolph Raff and Thomas Kaufman.

What literary character would you employ as a postdoc?

Candide. One always needs optimism.

What's your favourite conference destination, and why?

England. More relaxed than the United States, great historical settings, and better beer.

What was the worst/most memorable comment you ever received from a referee?

Several of our best-known papers were initially greeted with apathy or faint praise, or were not even reviewed upon submission. I won't say which ones in order to protect the innocent (and future submissions).

You have the audience in your hands, but some smart-alec asks you the killer question you have no idea how to answer. What's your response?

Tell them it's a great question and that I have no idea how to answer it. That usually leads

to more discussion and I get a better idea of the serious questions that other people have related to my discipline.

What book is currently on your bedside table?

In the Heart of the Sea: The Tragedy of the Whaleship Essex by Nathaniel Philbrick. Think lab science is a challenge? Try three months in a rowing boat in the open ocean with nothing to eat but the guy next to you.

What music heads the playlist in your car or lab?

Tom Petty and the Heartbreakers.

Assuming the dead can be raised and/or time travel exists, who from the world outside science would you most like to have dinner with?

Churchill, Roosevelt and Eisenhower in early 1944. I've read a fair amount by or about these figures and the history of the time. The weight and tension of that year, and all the issues of the next 50 years, would be a riveting discussion.

Where and when would you most like to have lived or worked?

Late nineteenth century in the American West, hunting dinosaurs.

You are on a plane behind two students obviously going to the same conference, who start to talk about your work. What do you do?

Take a nap.

What do you most dislike about having research published?

Reviewer roulette. Success is often more a matter of the luck of the draw of reviewers than of the diligence of authors. Referees and editors are crunched for time, so a piece of work that took years can be dismissed in minutes.

What would you have become, if not a scientist?

A (very) minor-league baseball player.

What music would you have played at your funeral?

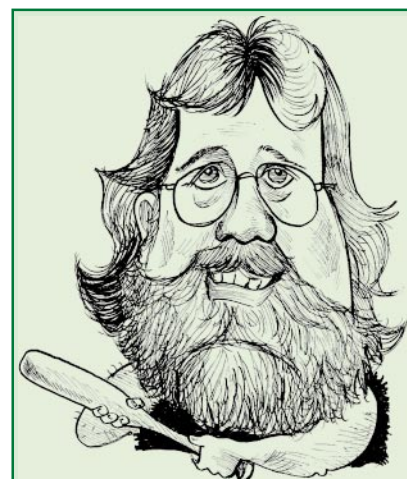
The Dixie Dregs — the best instrumental rock band, period.

The job of captain on the Starship Enterprise in Star Trek has become vacant. Nominate any real person, living or dead, for the post.

Robin Williams.

Is there a 'tyranny of reductionism' in how scientists are trained today?

Certainly it is a dominant viewpoint. Although I worry about the lack of natural history and organismal perspectives, the reductionist approach in biology has been spectacularly successful in addressing many complex mysteries for which it is appropriate.



Sean Carroll

Sean Carroll is an investigator of the Howard Hughes Medical Institute and professor of molecular biology and genetics at the University of Wisconsin-Madison. He studies animal development and evolution and occasionally co-produces rock videos depicting his lab's work.

Under what conditions do you have your greatest and most inspired ideas?

Sleeping, or near sleep.

What's the one thing about science that you wish the public understood better?

The depth and breadth of evidence supporting scientific ideas: compared with, say, the absence of evidence in areas such as astrology, UFOs and ghosts.

Which field in science (apart from your own) deserves more funding, and why?

Palaeontology. A lot of bang for the buck and great interdisciplinary training for future teachers and scholars.

What's the most interesting thing in your fridge?

Moët and Chandon Private Reserve, a souvenir of Epernay. Liquid gold.

Why is physics so hard?

After Newton's basics, I find that I have little intuitive feel for the rest.

What's just around the corner?

The rewriting of population genetics in terms of functional biology.

Which actor would best portray you in a film?

Jack Nicholson, directed by the Coen brothers.

You've just been told (in confidence) that the world will end tomorrow. What do you do next?

Party like ... there's no tomorrow.

What would be the title of your autobiography?

'Escape from Toledo'. ■

The supernova connection

Peter Mészáros

They are the most energetic events in the Universe, but the origin of γ -ray bursts has been hard to establish. Observations of a burst close to our Galaxy now show that supernovae are, as suspected, likely culprits.

The fog surrounding the identity of the progenitors of γ -ray bursts (GRBs) is beginning to lift, at least for the class of GRBs known as 'long' bursts. This is thanks to a series of observations of a burst that began on 29 March 2003, very close to our Galaxy. On pages 843, 844 and 847 of this issue, Uemura *et al.*¹, Price *et al.*² and Hjorth *et al.*³ reveal the evolution of this burst in unprecedented detail — and show that behind the GRB is the unmistakable signature of a supernova.

The GRB population divides neatly into long ones and short ones, depending on whether the burst of γ -rays lasts more or less than a few seconds⁴. About two-thirds of all observed bursts are long, and these are the only ones for which longer-lasting 'afterglows' at X-ray, optical and radio wavelengths have also been found. These afterglows may last up to several months, and from them the distance to the GRB and the identity of its host galaxy can be determined. There is good evidence that long bursts are largely associated with active, star-forming regions in small, blue galaxies. And, in at least three cases, there has been tantalizing evidence that GRBs are associated with a particular type of supernova⁴ — although that interpretation has so far been fraught with uncertainty.

A 'usual' supernova arises when the core of a massive star collapses, ejecting the stellar outer envelope. The majority of such supernovae result from parent stars that are less than about 30 times heavier than the Sun, and the core collapse produces a neutron star. These supernovae are normally detected weeks after the collapse, because the ejected envelope only brightens sufficiently to be detected at optical wavelengths some weeks later. The only signals of the collapse that are expected to reach the Earth promptly are a flux of tiny particles called neutrinos (which was picked up for the supernova SN1987a by the Japanese neutrino detector Kamio-kande) and gravitational waves (which have yet to be detected).

For heavier stars, however, the core is thought to collapse into a black hole, and the resulting brief episode of mass accretion has been proposed as the central engine driving GRBs⁵. This kind of collapse was initially referred to as a 'failed' supernova, as it was thought that the stellar envelope would not be ejected. A GRB would instead result from

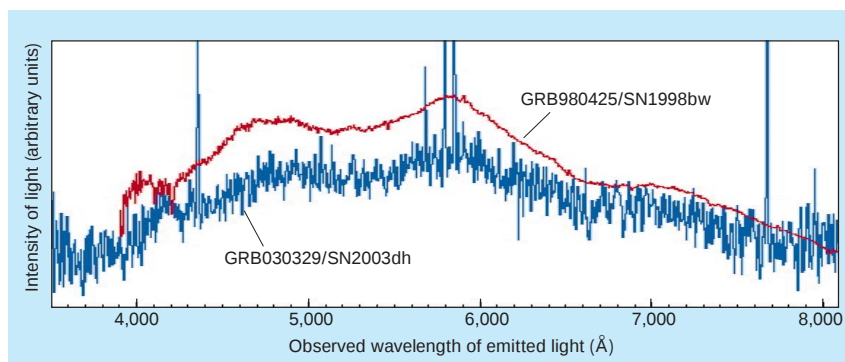


Figure 1 A good match. The spectra of the wavelengths of radiation from the γ -ray burst GRB030329, believed to be associated with the supernova SN2003dh, and from GRB980425/SN1998bw are remarkably similar in shape¹⁵, suggesting that in general the GRB and supernova phenomena are related. Detailed observations^{1–3} of GRB030329 offer the strongest proof yet that γ -ray bursts are indeed produced by supernovae that result when the core of a massive star collapses.

a relativistic jet of gas fed by the black hole; it would break through the stellar envelope, leading to radiative shocks in the rarefied environment outside the star.

In 1998, observations of GRB980425 showed an anomalous brightening of its optical afterglow a few weeks after the burst, possibly linking it to a roughly contemporaneous supernova, known as SN1998bw, whose ejected envelope would have brightened at about that time. Suspicions grew that long GRBs might, after all, be associated with 'successful' supernovae. In fact, the few supernovae tentatively linked to GRBs appeared even more energetic than usual, and were dubbed 'hypernovae'⁶, or 'collapsars'⁷. There is also a more elaborate offshoot of the supernova idea — the 'supranova'⁸. Here, the core collapse is assumed to be a two-step affair: the first step produces a temporary neutron star and a supernova; in the second step, a few weeks or months later, the neutron star collapses into a black hole, producing a GRB.

The association of long GRBs with supernovae (or even supranovae) is based on the approximate coincidence in time of the GRB and the inferred instant of the supernova core-collapse. The latter is deduced from an extrapolation back from the peak brightness of emitted light — an extrapolation that is model-dependent and uncertain, not least because these objects are so far away and very faint. This situation changed dramatically with the observation of GRB030329 in March of this year. Its coordinates were

disseminated by the HETE-2 spacecraft within 90 minutes of its detection, enabling ground-based telescopes to make follow-up observations almost immediately. Although more than two billion light years away, GRB030329 may be the nearest cosmological GRB yet seen^{3,9}. (In terms of the conventional astronomical distance measure, its 'red-shift', z , is 0.169; previous GRBs have usually only been seen in the range 0.4–4.5; the exception is GRB980425, if its association with SN1998bw at $z = 0.008$ is real.)

After a week, the pattern of light emitted by GRB030329 — its 'light curve' — started to show the beginnings of a slight bump. Ten days later, this bump was identified as being caused by an energetic supernova, labelled SN2003dh^{3,10}. Because this GRB is relatively close to us, the identification of the light-curve peak, and its wavelength spectrum, is significantly stronger than in previous cases. The extrapolation from the peak brightness indicates that the time off-set between the GRB and the collapse of the supernova is unlikely to be greater than about two days, and is compatible with the two events being simultaneous. Hjorth *et al.*³ interpret this as ruling out the supranova model: if the two-step collapse of a supranova were to happen within two days, it is unlikely that its γ -ray emission and afterglow would match the observations.

Another remarkable feature of GRB030329 is the spectrum of wavelengths it has emitted. There is a tantalizing match between the shape of the distribution for GRB030329/SN2003dh and that of

GRB980425/SN1998bw (Fig. 1) that signals a strong connection between the two phenomena. The light curve of GRB030329, as discussed by Price *et al.*², shows a steepening that they interpret as being caused by a broadening of the jet-like outflow from its progenitor. The light curve also contains a wealth of other, finely structured detail that has interesting implications for the identity of the central engine driving the GRB. Uemura *et al.*¹ have found numerous bumps in the optical light curve, whose relative duration remains approximately constant, and at each of which the normalization of the light curve moves up (Fig. 1 on page 843). Simple explanations — that a jet has encountered an inhomogeneous, blobby medium or is itself patchy — do not work. But there is a satisfactory explanation¹¹: as the leading edge of fast jet gas moves further away from the central engine of the burst, it starts to decelerate and is caught up by batches of slower-moving jet gas, creating a 'refreshed shock'¹². These slower batches of gas could have been ejected towards the end of the γ -ray-emitting period (which lasts tens of seconds); at each catch-up episode, more energy is added to the afterglow.

The detection of GRB030329 is a watershed event. It proves that some, if not all, long GRBs are definitely associated with core-collapse supernovae, and thus are a consequence of the evolution of massive stars. It also strengthens the interpretation that the variability of burst light curves over short timescales is caused (largely, if not exclusively) by variable input from the central engine, rather than by the γ -rays

encountering a variable external medium. These observations fill in important missing gaps in the phenomenology of GRBs.

But other questions have not been addressed, such as what the magnetic content of the jets¹³ is, or whether all long bursts are associated with supernovae. Also, what is the connection between GRBs and the related, but fainter and softer phenomenon of X-ray flashes? The identity of the progenitors of the short GRBs is also an open question (a candidate might be mergers of double neutron stars), but it may be addressed by the Swift mission¹⁴, to be launched this December. From the present observations^{1–3} and their interpretation, however, it is clear that a new plateau has been reached in our understanding of the GRB phenomenon. ■

Peter Mészáros is in the Departments of Astronomy and Astrophysics and of Physics, Pennsylvania State University, 525 Davey Laboratory, University Park, Pennsylvania 16802, USA.

e-mail: pmeszaros@astro.psu.edu

1. Uemura, M. *et al.* *Nature* **423**, 843–844 (2003).
2. Price, P. A. *et al.* *Nature* **423**, 844–847 (2003).
3. Hjorth, J. *et al.* *Nature* **423**, 847–850 (2003).
4. van Paradijs, J., Kouveliotou, C. & Wijers, R. A. M. *J. Annu. Rev. Astron. Astrophys.* **38**, 379–425 (2000).
5. Woosley, S. E. *Astrophys. J.* **405**, 273–277 (1993).
6. Paczynski, B. *Astrophys. J.* **494**, L45–L48 (1998).
7. MacFadyen, A. I. & Woosley, S. E. *Astrophys. J.* **524**, 262–289 (1999).
8. Vietri, M. & Stella, L. *Astrophys. J.* **527**, L43–L46 (1999).
9. Greiner, J. *et al.* *GCN Circ.* 2020 (2003).
10. Stanek, K. Z. *et al.* Preprint astro-ph/0304173 at <http://arXiv.org> (2003).
11. Granot, J., Nakar, E. & Piran, T. Preprint astro-ph/0304563 at <http://arXiv.org> (2003).
12. Rees, M. J. & Mészáros, P. *Astrophys. J.* **496**, L1–L4 (1998).
13. Coburn, W. & Boggs, S. E. *Nature* **423**, 415–417 (2003).
14. <http://swift.gsfc.nasa.gov>
15. Matheson, T. *et al.* *GCN Circ.* 2120 (2003).

aeons. Nowadays, there are relatively short regions at either end of the Y chromosome that are still identical to the corresponding regions of the X chromosome, reflecting the frequent exchange of DNA between these regions ('recombination') that occurs during sperm production⁴. But more than 95% of the modern-day Y chromosome is male-specific, consisting of some 23 million base pairs (Mb) of euchromatin — the part of our genome containing most of the genes — and a variable amount of heterochromatin, consisting of highly repetitive DNA and often dismissed as non-functional. Now, in an accomplishment that can only be described as heroic, Skalka *et al.*¹ report the complete sequence of the 23-Mb euchromatin segment, which they designate the MSY, for 'male-specific region of the Y'.

Prioritization in the Human Genome Project had led to the heterochromatic regions of the Y and other chromosomes being set aside to be dealt with later, if ever. But there was reason to hope that the euchromatin of the Y chromosome would present no more difficult a sequencing challenge than that found elsewhere in the genome. That supposition could not have been more wrong. As Skalka *et al.* report, the MSY is a mosaic of complex and inter-related sequences that made this one of the most problematic regions of the human genome thus far to be successfully sequenced and assembled.

For instance, about 10–15% of the MSY consists of stretches of sequence that moved there from the X chromosome within only the past few million years. These stretches are still 99% identical to their X-chromosome counterparts and are dominated by a high proportion of interspersed repetitive sequences, with only two genes. A further 20% of the MSY consists of a class of sequences ('X-degenerate' sequences¹) that are more distantly related to the X chromosome, reflecting their more ancient common origin. And the remainder comprises a web of Y-specific repetitive sequences that make up a series of palindromes — sequences that read the same on both strands of the DNA double helix, with two 'arms' stretching out from a central point of mirrored symmetry. These palindromes come in a range of sizes, up to almost 3 Mb in length, with more than 99.9% identity between the two arms of each palindrome.

The repetitive sequences, particularly the palindromes, caused some difficulties for sequence assemblers. Genome-sequencing projects involve fragmenting the genome in question into small, overlapping pieces, sequencing them, and then using computer algorithms to put the pieces together in the correct order. There are various ways of doing this; assembling the MSY's palindromes (and discriminating between their arms) required an iterative mapping and

Genome biology

Tales of the Y chromosome

Huntington F. Willard

Determining the sequence of the human Y chromosome presented a daunting challenge to genome researchers. But the task is now done, and the secrets revealed justify the effort.

Ancient maps showed the known world in colourful detail, beyond the edges of which lay vast expanses of *terra incognita*. Much creative thought went into portraying this unexplored territory, often featuring nasty-looking serpents and dragons. Only when Magellan managed to circumnavigate the globe did it become apparent that the unknown was in fact navigable, and that the serpents and dragons, if not illusory, could at least be tamed. The human genome has its *terra incognita* too, some of it known, much of it subject to alternating angst and fascination by genome biologists, and all of it to be avoided if possible — until now. On pages 825 and 873 of this issue^{1,2}, a group of modern-day Magellans describe how they sailed headlong

into the frothy seas of duplicated, inverted and otherwise troublesome sequences on the human Y chromosome. They have emerged safely on the other side, with tales to tell.

Because of its distinctive role in sex determination, the Y chromosome has long attracted special attention from geneticists, evolutionary biologists and even the lay public. It is known to consist of regions of DNA that show quite distinctive genetic behaviour and genomic characteristics. The two human sex chromosomes, X and Y (Fig. 1), originated a few hundred million years ago from the same ancestral autosome — a non-sex chromosome — during the evolution of sex determination³. They then diverged in sequence over the succeeding



Figure 1 Male make-up. The human X (left) and Y chromosomes, magnified about 10,000 times.

sequencing process more reminiscent of the knowledge-based mapping approaches of the early days of the Human Genome Project than the high-throughput assemblies that have emerged for most of the genome^{5,6}. This strategy was aided by the fact that the sequence came from a single Y chromosome, so Skaletsky *et al.* knew that minor sequence variations must have come from duplicated copies on the same chromosome, rather than from different Y chromosomes. Although necessarily more painstaking, this overall approach provides a model for how researchers might attack at least some of the troublesome areas of the rest of the genome — such as blocks of repetitive heterochromatin and the hundreds of regions of substantial sequence duplication⁷ — where standard assembly programs can be fooled.

This is not just a celebratory tale of a successful sequencing journey, however. Along the way, Skaletsky *et al.* picked up artefacts of Y-chromosome antiquity, dating as far back as 300 million years, that allow a glimpse into the evolutionary strategies that the Y chromosome has used to survive.

For instance, from the degree and patterns of divergence of the genes found on both sex chromosomes, the authors provide evidence for the stepwise decay of the Y chromosome over time and define changes in both Y-chromosome organization and gene content and expression. Unlike the regions at the ends, most of the lengths of the sex chromosomes do not exchange sequence during sperm production, and Skaletsky *et al.* point to two consequences of this suppression of recombination. First, selection occurred on the Y chromosome for a group of testis-specific genes that the authors argue may have enhanced male fertility. Most of these genes are found within the palindromes,

showing why it can be important to sequence such difficult regions. Second, as X–Y recombination became suppressed during evolution, an alternative mechanism had to emerge to maintain the sequence and function of the remaining Y-chromosome genes and to prevent the accumulation of inactivating mutations and the ultimate demise of the chromosome⁸.

To gain insight into this alternative mechanism, Rozen *et al.*² examined the hypothesis that X–Y recombination has been replaced by extensive, ongoing recombination between the arms of the MSY palindromes — where the sequence on one arm of the palindrome alters or ‘converts’ the sequence on the other. To test the predictions of this model, the authors sequenced one particular palindrome-embedded gene from Y chromosomes from around the world, representing the full tree of the previously established Y-chromosome genealogy⁹. They found several instances where the sequence of the copy of the gene on one arm of the palindrome had altered the sequence of the other arm’s copy. From this, they calculate that as many as 600 base pairs (from the 5.4 Mb contained in MSY palindromes) must be converted in each newborn male in the human population.

These data also indicate that gene conversion in general may be more common than previously suspected, especially in other palindromic and duplicated regions around the genome⁷. This supports a more dynamic view of genome change, in which, even within a single generation, not only does the occasional mutation occur (there are estimated to be as many as 100–200 new base-pair changes in each person), but also perhaps thousands of gene-conversion events.

The tales told by these Magellans of the genome hold two lessons for those who



100 YEARS AGO

May I record the discovery of musical sands at places along the shore between Ramsgate and Kingsgate. The sand occurs in small patches close to the chalk cliffs, the largest patch being found at Joss Gap. In composition the sand is very similar to that of Studland Bay, but the individual grains are more polished, and the proportion of denser materials far higher. Of course, the sand can only be experimented upon when it has been uncovered by the sea for a sufficient length of time to enable it to become dry, and it gives remarkable results when tested in the ordinary way — especially when placed in a china vessel and struck with a wooden plunger.

ALSO...

During a heavy thunderstorm at Heppner, Oregon, on Sunday last, a remarkable downpour of rain occurred, producing a destructive flood, which caused the death of more than three hundred people. Heppner is situated in a gulch through which a stream runs usually only a few feet in width. On Sunday a dense cloud suddenly covered the mountain overlooking the town, and the rain which followed produced a great mass of water which rushed down the mountain and carried everything before it... The flood swept a clean path more than a mile long and two blocks wide through the town.

From *Nature* 18 June 1903.

50 YEARS AGO

Interesting observations on the X-rays accompanying μ -meson capture by nuclei are reported through the rather unusual channel of a ‘press release’ from Columbia University. Negative μ -mesons brought to rest in matter eventually disappear, either by spontaneous decay or by nuclear capture... First the meson is attracted into quantum orbits around a nucleus, to which it is bound by the electrical forces. During this stage it emits X-rays or ejects atomic electrons as it jumps successively from one quantum state to another, until it reaches the state of lowest energy. Disintegration or capture then occurs... Rainwater and Fitch have observed these X-rays and made precise determinations of their energies. Since the deepest meson orbit and the nucleus are comparable in size for a heavy nucleus, the energy of the former is strongly influenced by the way in which electric charge is distributed in the nucleus... Rainwater and Fitch conclude that the charge is considerably more concentrated towards the centre than had been previously believed.

From *Nature* 20 June 1953.

might question the wisdom of such exploration. First, even the most repetitive and seemingly impenetrable stretches of the genome hold secrets that justify the effort. Second, each chromosome has its own story to tell, quite apart from the story of the genome as a whole. Although the sex chromosomes provide the strongest case for a special relationship between genome organization and the unique biology of a chromosome^{10,11}, the other chromosomes shouldn't feel left out. Each is the product of hundreds of millions of years of evolution, shaped by processes that have rearranged and exchanged sequences, contributed to the formation of new species, given birth to new genes and gene families, and provided the basis for a range of genetically determined or genomically influenced traits. Piecing together these events remains a

worthwhile challenge, for among the flotsam and jetsam of each chromosome lie clues to our history.

Huntington F. Willard is at the Institute for Genome Sciences and Policy, and the Department of Molecular Genetics and Microbiology, Duke University, Durham, North Carolina 27710, USA. e-mail: hunt.willard@duke.edu

1. Skaletsky, H. *et al.* *Nature* **423**, 825–837 (2003).
2. Rozen, S. *et al.* *Nature* **423**, 873–876 (2003).
3. Ohno, S. *Sex Chromosomes and Sex-Linked Genes* (Springer, Berlin, 1967).
4. Burgoyne, P. S. *Hum. Genet.* **61**, 85–90 (1982).
5. International Human Genome Sequencing Consortium *Nature* **409**, 860–921 (2001).
6. Venter, J. C. *et al.* *Science* **291**, 1304–1351 (2001).
7. Bailey, J. A. *et al.* *Science* **297**, 1003–1007 (2002).
8. Marshall Graves, J. A. *Trends Genet.* **18**, 259–264 (2002).
9. Cavalli-Sforza, L. L. & Feldman, M. W. *Nature Genet.* **33**, 266–275 (2003).
10. Lahn, B. T. & Page, D. C. *Science* **278**, 675–680 (1997).
11. Carrel, L., Cottle, A. A., Goglin, K. C. & Willard, H. F. *Proc. Natl Acad. Sci. USA* **96**, 14440–14444 (1999).

Global change

Ups and downs in the Red Sea

Frank Sirocko

Changes in past conditions in the Red Sea have been exploited to provide a detailed record of sea-level variation over much of the last glacial period. That record might tie in with events in the far south and north.

On page 853 of this issue, Siddall *et al.*¹ present a new approach to reconstructing the history of sea-level change during the last glacial cycle. They provide a detailed record, with a resolution on the century scale, spanning the years between 70,000 and 25,000 years ago. This was a time of intense overall glaciation, punctuated by abrupt warmings and coolings. The last glacial cycle culminated some 18,000 years

ago with the so-called Last Glacial Maximum, before leading into our present interglacial condition, beginning about 10,000 years ago. The reason why past sea levels are of such interest is that they reflect global temperatures — during colder episodes, more of Earth's water becomes locked up in ice caps, with a consequent fall in sea levels.

The authors' approach involved analysing oxygen-isotope values of the calcite tests of

foraminifera from Red Sea sediment cores. Foraminifera are planktonic organisms that originally lived in the surface waters, and accumulated calcite in equilibrium with an oxygen-isotope measure ($\delta^{18}\text{O}$) of the water in which they lived. The $\delta^{18}\text{O}$ composition of the Red Sea is unlike that of the world ocean, because in this semi-enclosed basin the water is subject to especially strong evaporation. Strong evaporation favours release of the lighter ^{16}O isotope over that of ^{18}O — the water is thus enriched in ^{18}O , and is more saline.

This process is also typical for some other ocean basins at low latitude, such as the Persian Gulf or the Mediterranean. But the Red Sea is unique in the characteristics of its connection to the open ocean, through the Gulf of Aden. This connection, the Strait of Bab el Mandab, is only 18 km wide. It also has a 'sill', which at present lies 137 m below the surface. During the 'low stand' of sea level at the Last Glacial Maximum, however, it was at a depth of only about 15 m (Fig. 1). Accordingly, during glacial times, the amount of surface water flowing into the Red Sea was usually very restricted. The residence time of water in the basin was therefore much longer than now, leading to an even greater evaporative effect on the $\delta^{18}\text{O}$ and salinity characteristics. Put another way, compared with that of other records, the sea-level signal seen in the Red Sea sediments is greatly amplified.

Siddall *et al.*¹ have combined a hydraulic control model² of water flow with an algorithm to calculate the equilibrium $\delta^{18}\text{O}$ values; these are a function of the composition of the inflowing water, the surface-water temperature, evaporation and the amount of inflow. The authors then applied this model to interpret the record of a sediment core, known as KL11, from the central Red Sea. The overall result is a history of sea level that shows several prolonged maxima and minima, especially between 70,000 and 40,000 years ago (see Fig. 2b, overleaf).

The broader significance of this work lies in how it might be related to the ice-core records at high latitudes, both south and north. Thus, the calculated salinities and model-derived sea levels show a succession of highs and lows that are similar to the temperature variations inferred from two Antarctic ice cores, Byrd and Vostok. The KL11 record shows for the first time that the temperature variations documented for the Antarctic were probably paralleled by changes in sea level (Fig. 2a, b).

A few years ago, the Byrd ice core was synchronized in time to the Greenland ice core, GISP2. To do this, the authors concerned³ matched variations in levels of atmospheric methane, which must have occurred simultaneously worldwide. After carrying out this exercise, their most astonishing observation was that climate change in the Antarctic apparently occurred several millennia before

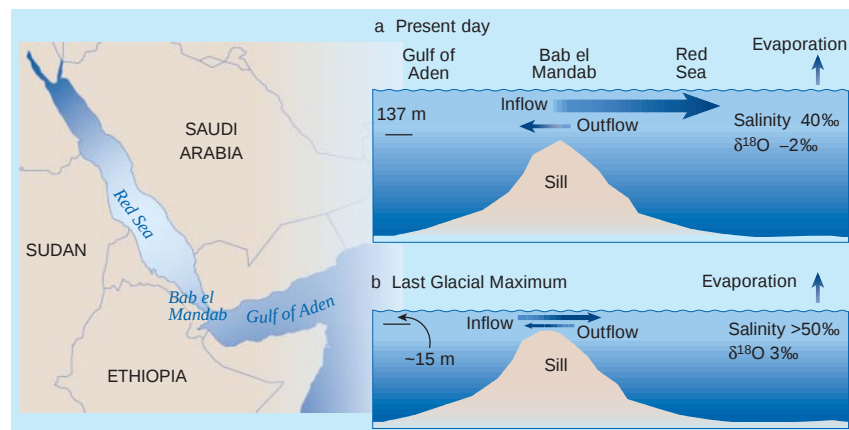


Figure 1 Now and then at the Strait of Bab el Mandab, which is connected to the Indian Ocean through the Gulf of Aden. **a**, Status at the present day, during an interglacial with a water depth of about 137 m at the sill. Typical water conditions are a salinity of 40 parts per thousand (‰) and a $\delta^{18}\text{O}$ value of -2‰ . In reality, the modern seasonal pattern of water flow is highly complex⁴. **b**, Status about 18,000 years ago, during a time of greatest cooling at the Last Glacial Maximum, with a water depth of approximately 15 m at the sill. The altered water-flow pattern and relatively stronger evaporative effect produced a salinity of more than 50‰, and a $\delta^{18}\text{O}$ value of 3‰. The size of the arrows is proportional to the strength of the processes indicated.

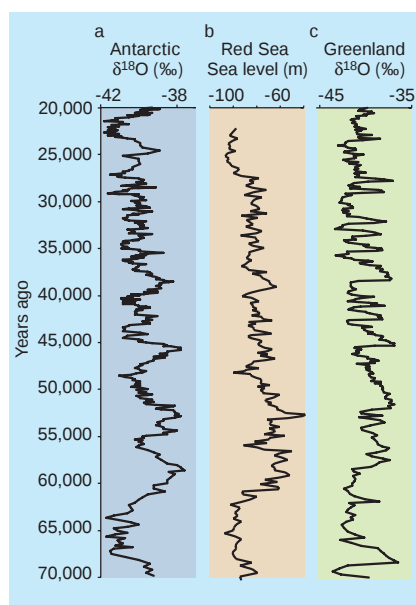


Figure 2 Comparison of records from the Antarctic, the Red Sea and Greenland. a, Data from the Byrd ice core from the Antarctic. These are $\delta^{18}\text{O}$ measurements in parts per thousand — the more negative the $\delta^{18}\text{O}$ figure, the cooler the climate. b, The Red Sea record of Siddall *et al.*¹, showing sea level in metres below today's level. c, Data from the GISP2 ice core from Greenland. These have the same units as a, with a and b having been brought into synchrony in earlier work³.

change in Greenland. If the sea-level record of Siddall *et al.* indeed follows that of temperature in the Antarctic, perhaps the chain of events was that Antarctic warming caused ice melting, and that the subsequent large rise in sea levels of several tens of metres in turn resulted in the collapse of continental ice sheets in the north. But however enticing this scheme may seem, there is an important caveat — in the Antarctic region itself, there

is no evidence as yet for melting of the Antarctic ice sheet during the interval of rapid climate changes between 70,000 and 25,000 years ago.

This view of events also assumes that the record inferred by Siddall *et al.* is robust — that is, for instance, that the $\delta^{18}\text{O}$ signal in the KL11 core has not been affected by environmental factors other than a sea-level-dependent basin-concentration effect, and that the Strait of Bab el Mandab has been stable over the period of time concerned. However, the range of 5‰ in the KL11 $\delta^{18}\text{O}$ curve (between -2 ‰ and 3 ‰; see Fig. 1) is a good indicator that the observed changes do indeed primarily represent the amplifying effect due to reduced exchange of water between the Red Sea and open ocean. Melting of the glacial continental ice would produce only a 1‰ change in the world ocean.

There are several methods for reconstructing sea levels, involving the measurement of oxygen-isotope variations of open-ocean water, lateral shifts in coastlines, and the height of shore terraces or drowned coral reefs. The beauty of Siddall and colleagues' approach compared with these other methods is that it can be applied to very high-resolution records as well as very long records. It promises to provide the data that will allow us to forge a firm link between events at low latitudes and those at high latitudes, and answer the fundamental question of whether the waxing and waning of the Antarctic ice sheet, and the consequent effects on sea level, have been the motor for abrupt climate change in the north. ■

Frank Sirocko is at the Institute for Geoscience, University of Mainz, 55099 Mainz, Germany.
e-mail: sirocko@mail.uni-mainz.de

1. Siddall, M. *et al.* *Nature* **423**, 853–858 (2003).
2. Siddall, M., Smeed, D. A., Matthies, S. & Rohling, E. J. *Deep Sea Res.* **149**, 1551–1569 (2002).
3. Blunier, T. & Brook, E. *Science* **291**, 109–112 (2001).
4. Ganssen, G. & Kroon, D. *Paleoceanography* **6**, 73–82 (1991).

Gene regulation

Versatile assembler

Steven Henikoff

The protein machines that assemble and remodel chromosomal proteins often contain a common component. The role of this component in maintaining stability during development is now revealed.

No handyman would be without WD-40, the penetrating oil that stops squeaks, loosens rusted parts and protects against corrosion. The impressive versatility of WD-40 puts it in a class with the Swiss army knife; however, unlike the knife, which sports a collection of specialized modules, the oil is applied in a similar way to perform each different task. An analogous situation appears to hold for a member of the WD-40 protein family. The protein in

question resembles its household namesake in displaying a surprising amount of versatility: it is found in several protein machines that apparently act in different ways on chromatin, a packaged form of DNA. Researchers studying this WD-40 protein suspect that it too, like the oil, functions in the same way to carry out these different tasks — but just how it works and why it is found in such diverse complexes is a mystery. Hennig and colleagues' study of a mutant plant, described

in a recent issue of *Development*¹, could provide a crucial clue.

WD-40 proteins are so called because they contain a repeated segment, of some 40 amino acids, that includes a tryptophan–aspartate peptide ('WD' in the single-letter code for amino acids). A set of seven tandem WD-40 repeats folds into a propeller shape, in which each propeller blade can serve as a platform for binding a different protein². For instance, the particular WD-40 protein studied by Hennig *et al.*¹ — RbAp48 — was first identified in mammals as a ubiquitous binding partner for the retinoblastoma protein, and so is presumably involved in cell-growth suppression like this protein³. Subsequently, a biochemical role for RbAp48 was revealed when it was found to be part of the human chromatin-assembly factor-1 (CAF-1) complex. Chromatin somewhat resembles a string of beads; each bead is known as a nucleosome particle and is composed of DNA that is tightly wrapped around an octamer of four types of histone protein. The CAF-1 complex promotes the assembly of nucleosomes onto newly replicated DNA (Fig. 1)⁴.

RbAp48 is far more abundant in the cell than the other two CAF-1 subunits, p150 and p60, hinting that it also contributes to other complexes. Indeed, several further protein machines that act on chromatin include RbAp48 as an integral component⁵. For example, the fruitfly *Drosophila melanogaster* has an RbAp48-like protein that was first found in a chromatin-remodelling complex called NURF — which moves nucleosomes along DNA⁶ — and was later discovered in other complexes that affect chromatin. Among these are several complexes containing the E(Z) and ESC proteins⁷, implicated in maintaining the silent state of master regulatory genes during fruitfly development. One specific role of these complexes is to help remove acetyl groups from histones H3 and H4 (ref. 7); another involves adding methyl groups to a lysine amino acid in histone H3 (ref. 8). So, fruitfly RbAp48 is found in complexes that contribute to nearly every stage of chromatin metabolism, from nucleosome assembly and remodelling to histone modification. Moreover, RbAp48 relatives in other species are involved in yet more chromatin-associated processes: one is part of the major yeast histone acetyltransferase enzyme⁹.

The number of processes involving RbAp48, and the fact that its encoding gene is found only once in the *Drosophila* genome, might explain why researchers have been unable to isolate or engineer fruitflies containing RbAp48 mutations: the animals simply cannot survive. That in turn means that the role of this protein during development has remained obscure. But the model plant *Arabidopsis* has five genes encoding

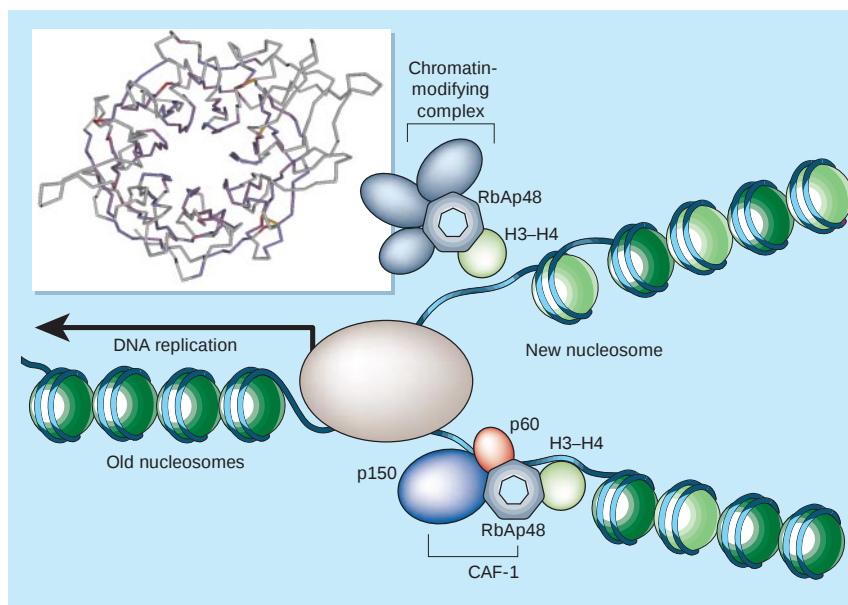


Figure 1 Chromatin assembly on new DNA. Chromatin consists of nucleosomes — DNA wrapped around a core of histone proteins — which are assembled anew as DNA is replicated during cell division. Old nucleosomes (dark green) are partitioned randomly to daughter DNAs at the replication 'fork'; new nucleosomes (light green) are assembled behind the DNA polymerization machinery (beige). Assembly begins with tetramers of histones H3 and H4 being deposited onto DNA; this is facilitated by CAF-1, which consists of p150, p60 and RbAp48 subunits. RbAp48 is part of the WD-40 family, the propeller-like basic structure of which is shown in the inset. RbAp48 works in CAF-1 by holding out histones for DNA to wrap around. It is also found in other chromatin-modifying complexes, and Hennig *et al.*¹ show that it is needed for passing on epigenetic information from cell to cell during *Arabidopsis* development. So, RbAp48 may be needed in a variety of complexes to hold out histones ready for modifying after DNA replication.

RbAp48-related proteins, which presumably have complementary or overlapping functions in chromatin metabolism; so mutating just one of these genes might be expected to have a relatively mild effect on development. Indeed, Hennig and colleagues' study¹ of an *Arabidopsis* RbAp48 mutation provides intriguing insights into the relationship between the protein and developmental gene silencing.

One of the RbAp48-like proteins in *Arabidopsis*, designated AtMSI1, appears to be the counterpart of mammalian RbAp48. Hennig *et al.* produced *Arabidopsis* mutants that possessed markedly reduced amounts of this protein. The mutant plants displayed a progressive failure of normal developmental gene silencing, with the floral master regulatory genes *AGAMOUS* and *APETALA2* showing unscheduled expression. Most

remarkable was the progressive nature of the resulting developmental defects. For example, the first flowers to form on a lateral branch were nearly normal, but as the branch grew and new flower buds appeared, they became increasingly affected, losing the normal morphology.

How can these progressive silencing defects be explained? Silencing is believed to be caused by the relative immobilization

Echolocation

Volume control

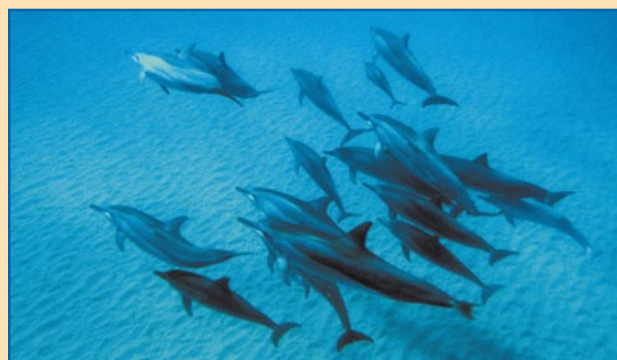
Bats, dolphins and submarines all use sound rather than vision to locate prey, friend or foe — bats because they hunt at night, dolphins and submarines because their marine environment is all but impenetrable to light. They emit sonar signals, and then process the echoes that are bounced back from distant objects to produce three-dimensional reconstructions of their targets. But there's a problem: as a target gets closer, fewer echoes are lost to the surrounding medium, so the sound should become louder. How can the sonar emitter avoid being deafened? Elsewhere in this issue (*Nature* **423**, 861–863; 2003), Whitlow W. L. Au and Kelly J. Benoit-Bird reveal how dolphins cope.

The authors used a symmetrical array of four hydrophones, submerged in the sea, to record the sounds emitted by several species of dolphin, including white-beaked dolphins, Hawaiian spinner dolphins

(pictured) and killer whales. The shape of the array was designed to ensure that the sounds it received were as close as possible in amplitude to the sounds at the centre (rather than the edges) of the beam of sonar 'clicks' emitted by the animals. The approach also allowed the authors to determine the distance of the animals from the array.

They found that, as the dolphins homed in on the hydrophone array, the amplitude of the sonar they emitted decreased by 6 decibels every time the distance was halved. That would ensure that the echoes do not increase in strength as the animals get closer to their target.

There's more than one way to solve this problem: bats keep the amplitude of their sonar signals constant, but decrease the sensitivity of their hearing once they have emitted the signals, allowing the sensitivity to increase gradually



with time. So, if the target is close by, the echo is heard at the same level as if it were far away. Submarine sonars use the same trick. Au and Benoit-Bird suggest that dolphins adjust their sonar transmitter, rather than their receiver, because the muscles in their middle ear are stiffer and denser than those of bats, making them less adaptable.

The authors also come up with a mechanism for how dolphins adjust the amplitude of the sonar clicks they produce. The rate at which clicks are emitted is known to rise as the animal closes in on its target. Au and Benoit-Bird speculate that dolphins decrease the amplitude of the clicks as the emission rate increases, thereby connecting amplitude to distance.

Amanda Tromans

of nucleosomes, which can present barriers to active processes such as gene expression⁵. This reduction in mobility might be achieved in many ways, including the deacetylation and methylation of histones, interactions with non-histone chromatin proteins, the use of variant histones, higher-order folding of chromatin, or a combination of these. Regardless of the precise mechanism, silencing must be maintained throughout many rounds of cell division during development — and it seems that RbAp48 is essential for this process. So, if RbAp48 levels are reduced, as in Hennig and colleagues' mutant plants, there is not enough protein to help keep every silent gene silent; some genes will become randomly activated, and will remain so during successive cell divisions. The progressive nature of the defects suggests that more and more genes become 'unsilenced' as development proceeds.

In other words, then, RbAp48 is part of a mechanism by which cells inherit 'epigenetic' changes — changes to chromatin that do not affect the DNA sequence. Just as DNA must be replicated during cell division, so too must patterns of methylation and acetylation when new nucleosomes are assembled. But how exactly does RbAp48 contribute? It is known that CAF-1 assembles nucleosomes onto newly synthesized DNA, and that the RbAp48 subunit acts as an 'adaptor' that attaches to tetramers of histones H3 and H4, making them available for the DNA to wrap around¹⁰. Specifically, RbAp48 attaches to a portion of H4 that becomes inaccessible once the nucleosome is assembled, suggesting that it can only act as an adaptor before this happens.

It is possible that RbAp48 has the same role in other chromatin-modifying complexes. CAF-1 is not essential in plants, animals and fungi¹¹, suggesting that other complexes can also assemble nucleosomes during DNA replication. Perhaps these other complexes have an RbAp48-like adaptor for holding onto the histone tetramer and presenting it for assembly (reviewed in ref. 12). And perhaps the other complexes that include RbAp48 and its relatives, such as histone-acetylation complexes, perform their tasks during replication-coupled nucleosome assembly — in which case, they might use RbAp48 to hold out exposed histones for modification. In this way, the diverse processes of histone modification and chromatin remodelling could contribute to epigenetic inheritance at the moment that histones are deposited onto newly synthesized DNA. Hennig and colleagues' finding¹ that RbAp48 is necessary for epigenetic inheritance lends weight to this idea.

If so, then this WD-40 protein might indeed resemble its namesake oil, acting in essentially the same way to perform a variety of tasks. ■

Steven Henikoff is at the Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA.
e-mail: steveh@fhcrc.org

1. Hennig, L., Taranto, P., Walser, M., Schönrock, N. & Gruissem, W. *Development* **130**, 2555–2565 (2003).
2. van der Voorn, L. & Ploegh, H. L. *FEBS Lett.* **307**, 131–134 (1992).
3. Qian, Y.-W. *et al. Nature* **364**, 648–652 (1993).
4. Verreault, A., Kaufman, P. D., Kobayashi, R. & Stillman, B. *Cell* **87**, 95–104 (1996).

5. Kadonaga, J. T. *Cell* **92**, 307–313 (1998).
6. Martínez-Balbas, M. A., Tsukiyama, T., Gdula, D. & Wu, C. *Proc. Natl Acad. Sci. USA* **95**, 132–137 (1998).
7. Furuyama, T., Tie, F. & Harte, P. J. *Genesis* **35**, 114–124 (2003).
8. Muller, J. *et al. Cell* **111**, 197–208 (2002).
9. Parthun, M. R., Widom, J. & Gottschling, D. E. *Cell* **87**, 85–94 (1996).
10. Verreault, A., Kaufman, P. D., Kobayashi, R. & Stillman, B. *Curr. Biol.* **8**, 96–108 (1998).
11. Kaya, H. *et al. Cell* **104**, 131–142 (2001).
12. Vermaak, D., Ahmad, K. & Henikoff, S. *Curr. Opin. Cell Biol.* **15**, 266–274 (2003).

Cosmology

Beyond the inflationary border

Steven Gratton and Paul Steinhardt

Observational data now offer strong support for inflation — a period of exponentially fast expansion in the early history of the Universe. But is the theory complete?

Cosmologists are exhilarated by recent cosmological observations¹ that have confirmed some of the long-standing predictions of the standard cosmological model, a picture based on a combination of Big Bang theory and inflationary cosmology. At the same time, some may worry that cosmology has “become a victim of its own success”² — that the basic outline of cosmic history is understood and all that remains is the unglamorous task of determining the precise quantitative details. But rest assured, there is much fundamental work still to be done, as a paper by Borde, Guth and Vilenkin³ in *Physical Review Letters* reminds us. The standard model is less a solid edifice than a scaffolding with many gaps, resting on uncertain foundations.

According to the standard picture, the Universe sprang into being about 14 billion years ago, as a space filled with matter and radiation of nearly infinite temperature and density. At birth, the space was likely to have been inhomogeneous, curved and warped, unlike the large-scale Universe we see today. After about 10^{-35} seconds, there began a brief period of exponentially fast expansion, known as inflation, that ironed out any curves or warps in space. Inflation also smoothed out the distribution of matter and radiation, leaving behind tiny wrinkles that match the observed spatial variations in the cosmic microwave background radiation and provide the seeds for galaxy formation.

The story has become familiar, but consider its foundations. Is there really a beginning to the Universe? What events led to the onset of inflation? And does the Universe even contain the ingredients necessary for inflation (in particular, the ‘inflaton field’ that purportedly drives inflation and then decays into hot matter and radiation)? Without answers to these questions, the model is incomplete. Most cosmologists have set these questions aside, assuming that advances in fundamental

physics (such as string theory) will address them. They have focused instead on the post-inflationary Universe and on comparing predictions with observations. Now that the comparison is proving to be a success, though, theorists will be giving increasing attention to the foundational issues.

A conceivable resolution is that the Universe was inflating for all time until 14 billion years ago, when the inflaton field decayed and produced a hot, expanding volume filled with matter and radiation that cooled to become our present Universe. There would be no beginning and no need for any special dynamics to kick-start inflation.

Not possible, say Borde, Guth and Vilenkin³. Given fairly mild assumptions, they track a flow of particles in a generic inflating universe and show that an extrapolation backwards in space and time must arrive at a border (a surface in space and time) beyond which there can be no inflation. So the question of what happened before inflation seems hard to avoid. (It's not impossible, though: an example has been constructed⁴ in which the original inflating region can be joined to a second one, but this is made possible only by imposing the conditions that time flows in opposite directions away from the border and that no particles or other information can pass through.)

One might think that the pre-inflationary phase is unimportant because inflation wipes out all of the details. But the border is only a finite distance away in proper time, so it is quite possible to imagine disturbances in the pre-inflationary phase that would survive to the present. In fact, only a subset of pre-inflationary conditions leads to a sufficiently homogeneous universe after inflation. So, to have a complete understanding of the present post-inflationary phase, we must understand what is on the other side of that border.

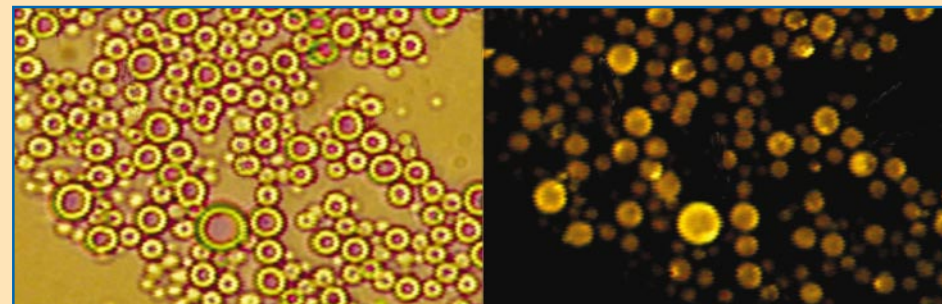
One possibility is a pre-inflationary phase that contracts for a semi-infinite period of

Materials science

Rings of excitement

Jennifer N. Cha and colleagues have taken a natural approach to create a new material for lasing applications. Emulating composite materials such as shell and bone, whose properties result from the cooperative organization of organic and inorganic components, Cha *et al.* have used specifically tailored polypeptides to organize inorganic quantum dots into microscopic laser cavities (*Nano Lett.* doi:10.1021/nl034206k; 2003).

The structures are created by mixing a solution of a positively charged diblock copolypeptide with a solution of negatively charged quantum dots — nanoscopic particles of a semiconducting material (in this case, cadmium selenide). Hollow, micrometre-size spheres form, which Cha *et al.* attribute to charge matching between the components of the mixture. The addition of silica nanoparticles to the mixture stabilizes the spheres, giving the double-layer structure of an inner wall of quantum dots and an outer



wall of silica nanoparticles that can be seen in optical-microscope images (such as the one shown here, left).

If the microspheres are excited by a laser pulse, the quantum dots emit light spontaneously; the spherical cavity acts as a resonator, modifying the resultant photoluminescence spectra. Although the microspheres are in contact with one another at their outer silica layers, the photoluminescence is confined to well-separated rings (right-hand

image): the light emitted by the quantum dots does not spread out, but is confined to the quantum-dot layer within the silica wall of each microsphere. If the intensity of the laser pulse is increased above a certain threshold value, a sharp peak appears in the photoluminescence spectra of each microsphere, and this increases rapidly with the energy of the laser pulse — a distinctive signature of lasing action in the microspheres.

Microspheres make room-

temperature microcavity lasing possible, without the need for mirrors or gratings. And, as the fabrication method is quite straightforward, it should be possible to tune the properties of both the quantum dots and the microcavity for particular lasing applications. The microspheres could also be assembled into arrays or more complex three-dimensional structures, their silica coating ensuring that individual spheres retain their functionality.

Magdalena Helmer

time and then bounces at the border into an inflationary phase. Another is a universe, expanding at a sub-inflationary pace, that joins onto an inflating phase. But because the inflaton field is unstable during inflation (it has to eventually decay into hot matter and radiation), it must also be unstable before inflation. It turns out that the probability of decay is exponentially greater in a contracting or slowly expanding phase than in an inflating one, so how could a sufficient region of the unstable inflaton field survive through the pre-inflationary phase?

Another option is to assume that quantum mechanics takes over before the border region is reached. The Hartle–Hawking ‘no boundary proposal’⁵ deals with the transition from quantum to classical cosmology, and many have hoped that this would naturally lead directly to a description of inflation on the classical side. Unfortunately, instead it leads typically to an almost empty universe in which little or no inflation occurred⁶.

The work of Borde *et al.*³, combined with these other attempts, forces us to realize that the inflationary story is still incomplete. And this is not the only unresolved issue. The model predicts the total energy density in the Universe correctly, but the nature of 96% of that energy is unknown⁷. Furthermore, despite two decades of studies, the fields

responsible for driving inflation have not been identified, and there is no accepted explanation for the finely tuned interactions that the fields must possess for inflation to end smoothly. So, there are good reasons to cheer the recent breakthroughs, but there are also many fundamental issues that remain to be explored. And there is perhaps even room for radical alternatives.

Steven Gratton and Paul Steinhardt are in the Department of Physics, Joseph Henry

Laboratories, Princeton University, Princeton, New Jersey 08544-0708, USA.

e-mails: sgratton@princeton.edu

steinh@princeton.edu

1. http://lambda.gsfc.nasa.gov/product/map/map_bibliography.cfm
2. Carroll, S. *Nature* **422**, 26–27 (2003).
3. Borde, A., Guth, A. H. & Vilenkin, A. *Phys. Rev. Lett.* **90**, 151301 (2003).
4. Aguirre, A. & Gratton, S. *Phys. Rev. D* **67**, 083515 (2003).
5. Hartle, J. B. & Hawking, S. W. *Phys. Rev. D* **28**, 2960–2975 (1983).
6. Hawking, S. W. & Turok, N. *Phys. Lett. B* **425**, 25–32 (1998).
7. Fukugita, M. *Nature* **422**, 489–491 (2003).

Cell cycle

Degradation ensures integrity

Anatoliy Li and J. Julian Blow

Protein degradation terminates a range of biochemical activities in living cells. New work shows that a component of the ‘licensing’ system for DNA replication must be degraded to prevent re-replication during cell division.

How would you ensure that a biochemical reaction is irreversible? This is a key problem in living organisms, because many of their control systems must progress unidirectionally through a series of different states. During the cell-division cycle, for instance, chromosomal DNA must first be duplicated (in the so-called S phase), and then the two copies must be segregated to daughter cells (during mitosis). Organisms

usually solve the problem of irreversibility by inactivating the proteins that catalyse each biochemical reaction after they have carried out their task. There are several ways of achieving this, but the most direct is simply to destroy the proteins. For example, it is known that the exit from mitosis is rendered irreversible by the degradation of mitotic proteins. Elsewhere in this issue (page 885), Zhong and colleagues¹ show that the ability

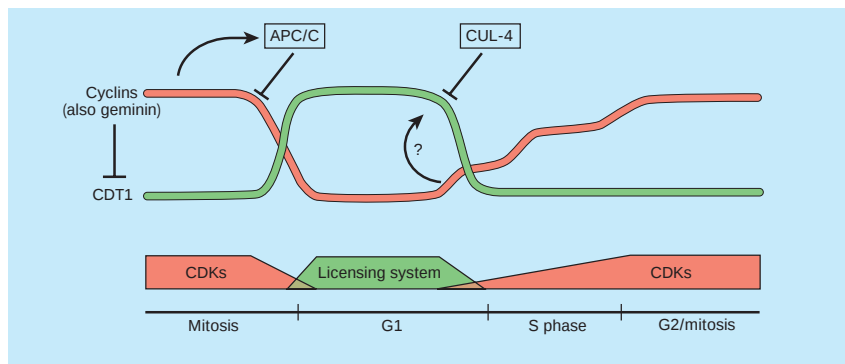


Figure 1 Protein degradation and the cell cycle. The four phases of the cell-division cycle are shown along the bottom; these are G1, S (when chromosomal DNA is precisely duplicated), G2 and mitosis (when the duplicated DNA is partitioned into two daughter cells). Cyclin-dependent kinases (CDKs) are activated during late G1 and are inactivated at the end of mitosis. The replication licensing system is active only in late mitosis and G1. The lines above show the relative abundances of cyclin proteins (required for CDK activity) and CDT-1 (essential for replication licensing). Cyclins are destroyed at the end of mitosis (along with the CDT-1 inhibitor geminin) by means of the anaphase-promoting complex (APC/C), a ubiquitin ligase; activation of the APC/C is dependent on CDKs. Zhong *et al.*¹ have now shown that CDT-1 is destroyed at the end of G1, and that its destruction requires the activity of the CUL-4 ubiquitin ligase. We speculate that CDT-1 becomes a CUL-4 substrate following phosphorylation by CDKs.

to precisely replicate chromosomal DNA also depends on protein degradation. Zhong *et al.* identify the protein CUL-4 as being essential for the timely degradation of another protein, CDT-1, which is a component of the cellular system that 'licenses' DNA replication. In the absence of CUL-4, cells continuously replicate their DNA without progressing into mitosis.

Proteolysis — the degradation of proteins into their constituent amino acids — is typically carried out by a cellular machine known as the 26S proteasome. Proteins are targeted to this machine by being covalently tagged with a short peptide called ubiquitin. So, the regulation of proteolysis depends on the activity of the ubiquitin ligase enzymes that attach ubiquitin to doomed proteins. One such enzyme is the anaphase-promoting complex, which is activated at the end of mitosis and, by ubiquitinating key proteins such as the cyclins, enforces mitotic exit (Fig. 1). CUL-4, a member of the cullin family of proteins, is an essential part of another ubiquitin ligase². Zhong *et al.*¹ have now found that, during larval development of the nematode worm *Caenorhabditis elegans*, the *cul-4* gene is expressed at high levels in proliferating cells. This hints that CUL-4 is involved in regulating cell-cycle progression — and so what, if anything, is its role?

To find out, Zhong *et al.* used RNA interference technology to reduce *cul-4* expression in *C. elegans* larvae. This caused several cellular abnormalities, the most dramatic of which was massive re-replication of DNA, generating cells with more than 30 times the normal DNA content. Cells normally use a strict control system to ensure that DNA replicates only once in each cell cycle; for this, replication is divided into two

non-overlapping phases³. Prior to S phase, replication origins (the places on DNA where replication starts) are 'licensed' by being loaded with the MCM2–7 protein complex. Licensing also requires several other proteins, including CDT-1. Then, during S phase, the replication machinery is assembled only at these licensed replication origins. As the DNA is replicated, the MCM2–7 proteins are displaced from it, thereby ensuring that replicated DNA is unlicensed and cannot be replicated again.

In order for this to work properly, it is essential that the licensing system is inactivated once S phase has started. Consistent with results in other organisms^{4–6}, Zhong *et al.* show that *C. elegans* cells contain CDT-1 only in late mitosis and G1 phase (the phase between mitosis and S phase; Fig. 1). Moreover, the authors find that reducing *cul-4* expression stabilizes CDT-1 in S phase, showing that CUL-4-mediated proteolysis is essential for the normal downregulation of CDT-1.

Is this stabilization of CDT-1 enough to explain why cells lacking CUL-4 over-replicate their DNA? Zhong *et al.* first provide several lines of evidence to support the idea that the over-replication is due to 'origin re-firing' — the continuing re-licensing and reactivation of replication origins during S phase³. They go on to show that, in strains heterozygous for *cdt-1* (that is, strains that have just one copy of the *cdt-1* gene rather than the normal two, and which therefore would be expected to have lowered CDT-1 levels), the re-replication induced by the reduction in CUL-4 levels is decreased fivefold. In contrast, strains heterozygous for the gene encoding another licensing protein, *cdc-6*, show no major reduction in re-replication when CUL-4 levels are decreased.

These results suggest that CUL-4-mediated degradation of CDT-1 is essential for the inactivation of the licensing system in *C. elegans*. Further evidence that levels of CDT-1 are critical for preventing DNA re-replication comes from a recent study of human cancer cells, where artificial over-expression of this protein was enough to cause significant re-replication⁷. Yet it is not known exactly how these levels are achieved — that is, how the activity of CUL-4 is regulated. For other cullin-based ubiquitin ligases, the target protein must be phosphorylated before it can be ubiquitinated. Given that cyclin-dependent kinases (CDKs) are known to be involved in preventing re-replication of DNA³, a plausible hypothesis is that, as CDKs are activated at the onset of S phase, they phosphorylate CDT-1, thus directing it to CUL-4 for ubiquitination.

Previous studies in other organisms have identified several ways in which the licensing system can be downregulated in S phase. These include inhibitory phosphorylation of licensing components by CDKs, and the inhibition of CDT-1 by a protein called geminin, as well as proteolysis³. In budding yeast, for example, three different regulatory mechanisms must be inactivated to promote modest re-replication of DNA⁸. One striking feature of Zhong and colleagues' work is the apparent dependence of *C. elegans* cells exclusively on CUL-4-mediated degradation of CDT-1 — in the absence of which massive re-replication occurs.

Most cancer cells display genomic instability, and often also dysregulated ubiquitin ligases, leading to the inappropriate 'survival' of their protein targets⁹. Significantly, it has been found that *cdt-1* can potentially cause cells to become cancerous¹⁰, and, in light of the new work^{1,7}, we speculate that this results from aberrantly expressed *cdt-1* driving over-replication, thereby causing genomic instability. The extensive re-replication that occurs with the loss of CUL-4 function in *C. elegans* suggests that, in some cases, the barriers to genomic instability — which in multicellular organisms can potentially lead to death from cancer — are less robust than might be imagined.

Anatoliy Li and J. Julian Blow are in the Cancer Research UK Chromosome Replication Research Group, Wellcome Trust Biocentre, University of Dundee, Dundee DD1 5EH, UK.
e-mail: j.j.blow@dundee.ac.uk

1. Zhong, W., Feng, H., Santiago, F. E. & Kipreos, E. T. *Nature* **423**, 885–889 (2003).
2. Zheng, N. *et al.* *Nature* **416**, 703–709 (2002).
3. Blow, J. J. & Hodgson, B. *Trends Cell Biol.* **12**, 72–78 (2002).
4. Nishitani, H., Lygerou, Z., Nishimoto, T. & Nurse, P. *Nature* **404**, 625–628 (2000).
5. Wohlschlegel, J. A. *et al.* *Science* **290**, 2309–2312 (2000).
6. Nishitani, H., Taraviras, S., Lygerou, Z. & Nishimoto, T. *J. Biol. Chem.* **276**, 44905–44911 (2001).
7. Vaziri, C. *et al.* *Mol. Cell* **11**, 997–1008 (2003).
8. Nguyen, V. Q., Co, C. & Li, J. J. *Nature* **411**, 1068–1073 (2001).
9. Spruck, C. H. & Strohmaier, H. M. *Cell Cycle* **1**, 250–254 (2002).
10. Arentson, E. *et al.* *Oncogene* **21**, 1150–1158 (2002).

Neurobiology

Muscle molecule in myelin moderation

Neuron **38**, 747–758 (2003)

The dystroglycan protein is a well-known component of skeletal muscle, where it forms part of a complex that anchors muscle cells to the surrounding environment. Abnormalities in this complex can lead to the muscle-weakening disease muscular dystrophy. But dystroglycan is also found in non-muscle tissues, where it forms a similar complex — and where its function is less certain. Fumiaki Saito and colleagues now propose a role for the protein in the peripheral nervous system.

The authors produced mice that lacked the protein specifically in their Schwann cells. Normally, these cells wrap peripheral nerves in a fatty sheath of myelin, helping them to conduct electricity efficiently. But in the mutant mice, the dystroglycan complex was disrupted, the myelin sheath was abnormally folded, and nerves conducted electricity more slowly. There were also fewer voltage-gated sodium channels in the exposed nodes between adjacent Schwann cells; normally, these channels control the entry of sodium ions into nerve axons, helping action potentials to spread. Over time, the mice developed balance problems, and in older animals some axons were lost altogether.

Saito *et al.* conclude that dystroglycan is important in myelination and nodal structure in the peripheral nervous system, although quite how it works remains unknown.

Helen R. Pilcher

Ecology

Study in blue

Proc. R. Soc. Lond. B doi:10.098/rspb.2003.2390 (2003)

The plumage displays of some birds depend on 'structural colour', an optical effect caused by anatomical features of their feathers. Matthew D. Shawkey and colleagues have used a combination of spectrometry and transmission electron microscopy to study the ultraviolet–blue colour of male eastern bluebirds (*Sialia sialis*, pictured) in the United States, as produced in particular by the barbs of rump feathers which contain a 'spongy' layer of keratin rods and air spaces.

One conclusion that the authors reach is that colour production in this species is due to coherent scattering of light by the spongy layer. That is, they find that the observed colour is a result of the overall effect of scattering, rather than the preferential transmission of shorter wavelengths.

More notably, they believe they have identified a mechanism for variation in



ultraviolet–blue coloration between individual males. The biological significance of this lies in the possible use of colour as an indicator of male quality by females assessing potential mates. Shawkey *et al.* find that variation in the diameter of keratin rods is negatively correlated with 'spectral saturation'. This is the extent to which a certain wavelength predominates in the reflected colour and thus is a measure of colour purity.

Michael Hopkin

Oceanography

Fresh input

Geophys. Res. Lett. doi:10.1029/2003GL017065 (2003)

From time to time, especially large bursts of fresh water enter the far North Atlantic in places such as the Labrador Sea between Canada and the southern tip of Greenland. These 'great salinity anomalies' feed into the overall ocean circulation pattern that includes the Gulf Stream, and might influence its stability. To investigate the origin of these anomalies, H. Haak and colleagues have analysed the formation and thickness of sea ice in the Arctic Ocean over the period 1948–2001, its movement southwards, and the consequent pulses of fresh water entering the Labrador Sea.

They conclude that the main drivers of large salinity anomalies are events in the Arctic, with maxima in sea-ice creation being propagated across the North Pole and through the Fram Strait east of Greenland. Conditions in the Labrador Sea itself, or ice and freshwater movement through the Canadian Archipelago to the west of Greenland, seem to be of minor importance in most cases. The simulations show little overall effect of individual salinity anomalies on larger-scale circulation in the North Atlantic. But much might depend on the frequency of anomalies or the impact of particular atmospheric conditions, especially wind patterns.

Tim Lincoln

Chemistry

Krypton goes organic

J. Am. Chem. Soc. **125**, 6876–6877 (2003)

It has been recognized since the 1960s that the noble gases are not really so noble. Various compounds of these 'inert' elements have been made, including the first xenon-containing molecule in 1962 and a

compound of argon in 2000. Leonid Khriachtchev and colleagues have now introduced the noble gases into organic chemistry. They have made a compound in which krypton is inserted into acetylene (C_2H_2) to make a chemical bond between carbon and krypton.

This is not wholly unprecedented: Khriachtchev *et al.* have themselves previously forged a bond between krypton and cyanide, and the ionic species CH_3Kr^+ has also been reported. But $HKrCCH$ is arguably the first truly 'organic' neutral krypton molecule.

It is made by using ultraviolet light to photolyse acetylene trapped in a solid krypton matrix, and then annealing at 30 K. The new molecule can be seen through infrared spectroscopy. The researchers' quantum-mechanical calculations suggest that the Kr–C bond is mostly ionic, with a covalent contribution of about 20%. They say that this synthesis could provide a new point of entry into krypton chemistry, which has previously stemmed from KrF_2 . But that may depend on the ability to make $HKrCCH$ in more than trace amounts.

Philip Ball

Plant biology

Growing pains

Development **130**, 3137–3146 (2003)

Cells of the thale cress *Arabidopsis* are normally neatly shaped, but in a particular *Arabidopsis* mutant, called *crooked*, they look somewhat dishevelled. Jaideep Mathur and colleagues suggest that a jammed cytoskeleton underlies this shabby appearance.

To generate a particular cell shape, specific places on the outer membrane (and cell wall in plants) need to expand more than others. This is known to require the cytoskeleton component actin. Mathur *et al.* now note that *crooked* cells contain unusual, dense accumulations of actin — particularly after a period of rapid growth. And expansion takes place only in those areas that are served by a finer actin mesh. They find that the *crooked* cell shape results from a mutation that disables ARPC5, a protein involved in weaving fine actin meshes, and that expression of intact ARPC5 forces the cells back to their normal shape. Even human ARPC5 has this effect — showing that regulators of actin have been conserved during evolution.

Mathur *et al.* propose that actin filaments can form a fine mesh that allows vesicles carrying membrane or wall components to travel to the surface, or a dense, tight net that traps them. Thus, the actin density determines when and where vesicles deliver their cargo and allow membrane expansion, and so may account for the different cell shapes seen in plants.

Marie-Thérèse Heemels

The missing link in *Ginkgo* evolution

The modern maidenhair tree has barely changed since the days of the dinosaurs.

The maidenhair tree, or *Ginkgo*, is a gymnosperm that has been described as a 'living fossil' because it is known to have existed early in the Jurassic period 170 million years (Myr) ago, but a full understanding of its evolution has been impeded by a gap in the fossil record of more than 100 Myr — a crucial period during which the modern ovulate organs evolved from the Jurassic type¹. Here we describe a new *Ginkgo* fossil that was collected from the Lower Cretaceous fossil Lagerstätte (the Yixian Formation², which is over 121 Myr old³) in China and which fills this gap. This missing link reveals that *Ginkgo*'s reproductive structures at that time were more like those of the present-day *Ginkgo biloba* than those of the primitive Jurassic type, indicating that their morphology has changed little for over 100 Myr.

The fossil was found on the southern slope of the Yinwoshan Mountain in Toudaohezi village (41° 46' N, 121° 40' E) of Yixian County in Liaoning Province. The ovulate organs consist of clusters of collared ovules and a common stalk (peduncle) (Fig. 1a, c, d) that is 32–42 mm long and 1–3 mm wide. Only one to three large,

probably mature ovules are still attached (Fig. 1c, d), but judging by the number of collars, the organs originally had more (up to six) ovules. Each ovule is borne at the end of an individual stalk (pedicel) (Fig. 1a) that is about 2–3 mm in length and 2 mm in width in juvenile organs. In mature organs, the individual stalk is hardly visible below the collar (Fig. 1c, d). Well-developed ovules (seeds) are roughly circular in outline (7.3–8.8 mm long and 6–8 mm wide). The surface of the ovule is smooth.

The associated leaves (Fig. 1b, and see supplementary information) are small, consisting of a petiole and a flabellate-to-semicircular lamina that are dichotomously divided several times. The ultimate segments are wedge-shaped to oblanceolate, with an obtusely rounded or truncated apex and three to six veins that converge slightly towards the apex.

No record existed of well-preserved *Ginkgo* ovulate organs from between the Middle Jurassic and the Early Tertiary epoch⁴ (see supplementary information). The new Cretaceous *Ginkgo* is unlike the oldest known (Middle Jurassic) species *G. yimaensis*¹ (Fig. 1e), which has ovulate

organs with three or four mature ovules, each terminating in a long individual stalk. It also differs from the Early Tertiary (about 56 Myr ago) *G. adiantoides*⁵ and the living species *G. biloba*. The ovulate organs of both have only a single, developed ovule and an aborted one attached to the common stalk (Fig. 1e; for a comparison of the ovulate organs of living and fossil *Ginkgo* species, see supplementary information). Two or more ovules, each with an individual stalk, can occur in juvenile or aberrant organs of *G. biloba*^{6,7}, but these organs are usually shed in the early stages of development.

The associated leaves are not divided to the same extent as those of Jurassic species such as *G. yimaensis*¹ and *G. huttonii* (Sternberg) Heer⁸, but are more divided than leaves of the Early Tertiary and modern species (although leaves from new long shoots or seedlings of *G. biloba* are also deeply divided⁹) (Fig. 1e).

The new *Ginkgo* is morphologically intermediate between the Jurassic species and the Early Tertiary and modern species, and is closer to the latter types. The new finding extends the geological range of

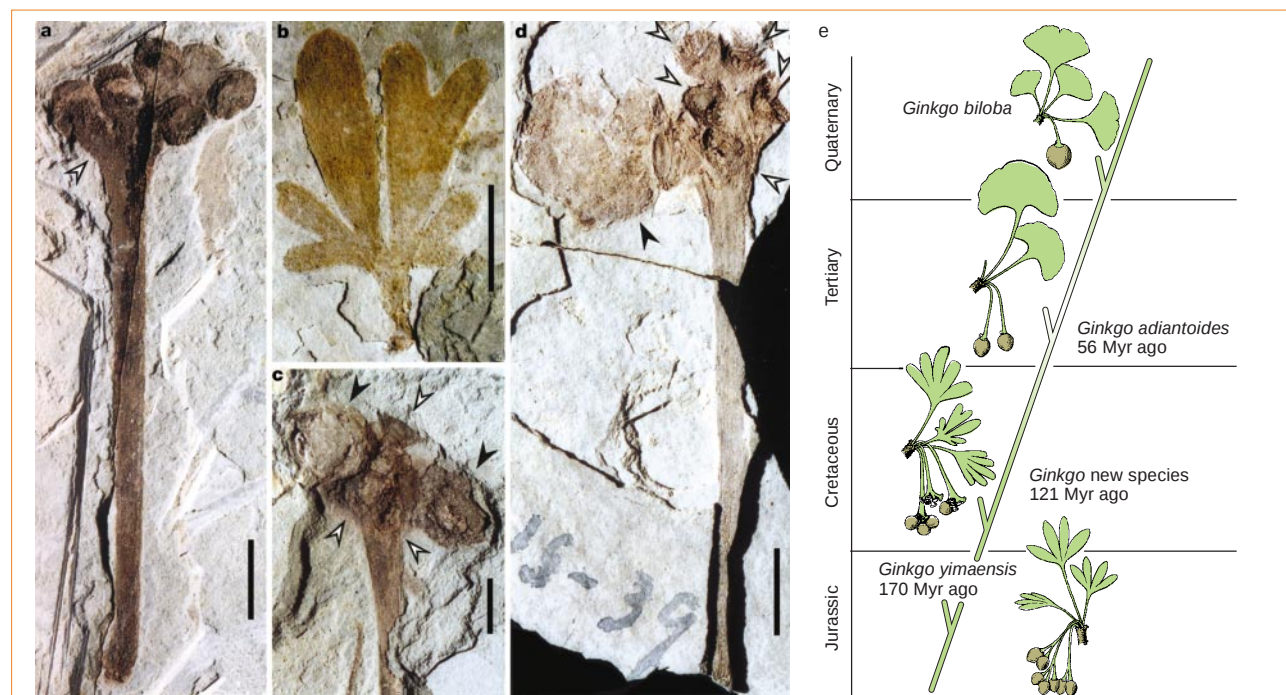


Figure 1 Newly discovered *Ginkgo* species from the Lower Cretaceous Zhuanchengzhi Bed of the Yixian Formation, China. **a**, Juvenile ovulate organ with six collars and very short individual stalks (arrowed) (specimen YWS139, PB19880). **b**, An associated leaf (specimen YWS129, PB19881). **c**, Ovulate organ bearing two developed ovules (black arrows) and three empty collars (white arrows) (specimen YWS42, PB19883). **d**, A mature ovulate organ without individual stalks, bearing one developed ovule (black arrow) and five empty collars (white arrows) on the common stalk (specimen YWS139, PB19884). Scale bars, 5 mm. **e**, Evolution of the *Ginkgo* genus in geological history, showing a reduction of individual stalks and a decrease in the number of ovules, and an increase in the size of the ovules and in the width of leaf segments. These evolutionary trends are roughly consistent with the ontogenetic sequence of the living species and are probably caused by peramorphosis^{7,10}. Note that the modern type of ovulate organ, the ovule of which has no individual stalk when mature, appeared early in the Lower Cretaceous epoch.

the modern form and is evidence of a roughly 120-Myr morphological stasis in ovulate organs of *Ginkgo* (Fig. 1e). It also suggests that ovulate organs of the *G. biloba* type could have originated by heterochrony (peramorphosis)¹⁰ from the Jurassic *G. yimaensis* type⁷.

Zhiyan Zhou*, Shaolin Zheng†

*Nanjing Institute of Geology and Palaeontology, Chinese Academy of Sciences, Nanjing 210008, China
e-mail: zyzhou@nigpas.ac.cn

†Shenyang Institute of Geology and Mineral Resources, Ministry of National Land and Resources, Shenyang 110032, China

1. Zhou, Z.-Y. & Zhang, B.-L. *Palaeontographica B* **211**, 113–133 (1989).

2. Chang, M.-M., Chen, P.-J., Wang, Y.-Q. & Wang, Y. (eds) *The Jehol Biota* (Shanghai Science & Technology, Shanghai, 2001).
3. Swisher, C. C. III et al. *Chinese Sci. Bull.* **47**, 135–138 (2002).
4. Rothwell, G. W. & Holt, B. in *Ginkgo biloba — A Global Treasure from Biology to Medicine* (eds Hori, T. et al.) 223–230 (Springer, Tokyo, 1997).
5. Crane, P. R., Manchester, S. R. & Dilcher, D. L. *Fieldiana (Geol.) N. Ser.* **20**, 1–63 (1990).
6. Fuji, K. *Bot. Mag. (Tokyo)* **10**, 104–110 (1896).
7. Zhou, Z.-Y. *Acta Palaeontol. Sin.* **33**, 131–139 (1994).
8. Harris, T. M., Millington, W. & Miller, J. *The Yorkshire Jurassic Flora IV: Ginkgoales and Czekanowskiales* (Br. Mus. Nat. Hist., London, 1974).
9. Seward, A. C. *Fossil Plants Vol. 4* (Cambridge Univ. Press, Cambridge, 1919).
10. Alberch, P., Gould, S. J., Oster, G. F. & Wade, D. B. *Paleobiology* **5**, 296–317 (1979).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

Ion channels

Opposite thermosensor in fruitfly and mouse

Several members of the TRP (for transient receptor potential) family of ion channels act as physiological temperature sensors in mammals^{1–6}, but it is not known whether the invertebrate TRP subfamilies that are found in the fruitfly *Drosophila* and the roundworm *Caenorhabditis elegans* can be directly activated by temperature. Here we show that the *Drosophila* orthologue of ANKTM1, which is a cold-activated ion channel in mammals, responds to a warming rather than a cooling stimulus. The thermosensing function of these channels is therefore evolutionarily conserved, and they show a surprising flexibility in their response to different temperature ranges.

In mammals, four TRPVs (members of the vanilloid subfamily of TRP channels) are activated at distinct heat thresholds (33–52 °C), whereas TRPM8 (of the melastatin subfamily) and ANKTM1 are activated at cold (17–25 °C) temperatures. However, the molecular mechanisms that underlie thermal preference in *Drosophila* are not well understood^{7,8}. To identify potential invertebrate temperature-activated ion channels, we analysed the sequences of predicted TRP channels from *Drosophila* and *C. elegans*, focusing on orthologues of the mammalian thermosensitive TRPs. We operationally defined orthologues as reciprocal best BLAST hits on a comparison of the two genomes, which might also be indicative of a common ancestry. By this definition, mammalian TRPM8 and TRPV1–4 channels do not have invertebrate orthologues (results not shown).

Consistent with this, the *C. elegans* TRPV homologues, which have diverse sensory functions, are not directly activated by temperature⁹. Mammalian ANKTM1 belongs to a branch of TRP channels that includes four *Drosophila* and two *C. elegans*

predicted proteins (Fig. 1a). One of these proteins, Painless, is required for response to noxious thermal and mechanical stimuli in *Drosophila* larvae, although it is unclear whether this channel is a direct sensor¹⁰.

Another of these relatives is a sequence orthologue of mouse ANKTM1 (mANKTM1; GenBank accession number AY231177); we term this *Drosophila* ion channel dANKTM1 (GenBank accession number CG5751). It is more similar to mANKTM1 (32% identical, 54% similar) than to Painless (GenBank accession number CG15860; 22% identical, 39% similar) and has 13 predicted amino-terminal ankyrins and six transmembrane domains. We amplified a full-length, 3.5-kilobase dANKTM1

complementary DNA from adult *Drosophila* RNA by using the polymerase chain reaction with reverse transcription (sequence deposited under GenBank accession number AY302598), and analysed its behaviour as an ion channel.

Cooling temperature steps did not elicit currents at –70 mV from oocytes expressing dANKTM1, but they elicited strong inward currents in those that expressed mANKTM1 and human ANKTM1 (Fig. 1b, and data not shown). However, transient currents were consistently activated in response to warming in oocytes expressing dANKTM1, with a threshold of 24–29 °C ($n = 18$) (Fig. 1c). The heat-activated dANKTM1 current was outwardly rectifying and reversed near –30 mV (Fig. 1d), indicating that dANKTM1 is a relatively non-selective cation channel, as is mANKTM1 (ref. 6). In calcium-imaging experiments, dANKTM1-transfected Chinese hamster ovary cells also responded to a warming stimulus, with an activation threshold of about 27 °C (results not shown)⁶.

To our knowledge, this is the first characterization of an invertebrate temperature-activated ion channel. Given that *Drosophila* strongly prefers a temperature of 24 °C (ref. 7), the activation threshold of dANKTM1 at about 24–29 °C suggests that this ion channel might have a physiological role in heat sensing. Although mANKTM1 and dANKTM1 are sequence orthologues, they do not seem to be functional orthologues — the former

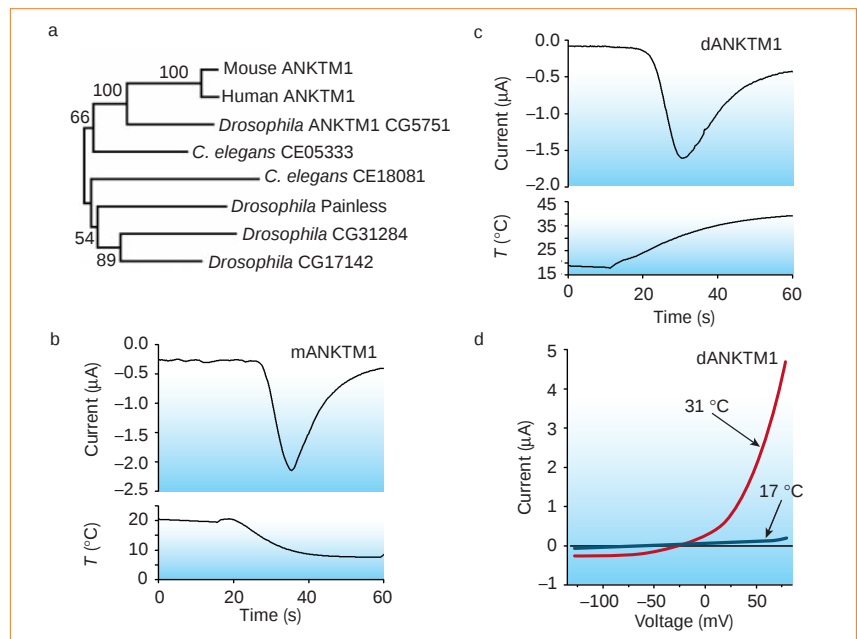


Figure 1 *Drosophila* orthologue of the mammalian ion channel ANKTM1 is activated by warm, not cold, temperatures. **a**, Phylogeny of human, mouse, *Drosophila* and *C. elegans* members of the ANKTM1 gene family. Branch lengths reflect the degree of divergence and the numbers indicate the percentage replication of branch points in bootstrap tests. NOMPC, a TRP channel involved in *Drosophila* mechanosensation¹¹, defines a separate subfamily and so is not included here. **b–d**, Temperature activation of mouse (**b**) and *Drosophila* (**c**, **d**) ANKTM1 in transfected *Xenopus* oocytes. Whole-cell currents were recorded at –70 mV in response to cooling and warming steps. Currents recorded in response to 2-s voltage ramps from –130 to 80 mV are shown in **d**. For mouse ANKTM1, warming from 20 to 40 °C resulted in almost no change in current, whereas *Drosophila* ANKTM1 produced no current increase in response to cooling from 20 to 7 °C (data not shown).

senses cooling, whereas the latter senses warming. The two proteins are 54% similar throughout their length, so it is not obvious which domains are crucial for the warm or cold response. But analysis of chimaeric mouse and *Drosophila* ANKTM1 proteins should help in the mapping of temperature-activation domains of these TRP channels.

Veena Viswanath*, Gina M. Story†, Andrea M. Peier*, Matt J. Petrus*, Van M. Lee*, Sun Wook Hwang†, Ardem Patapoutian*†, Tim Jegla*

*Genomics Institute of the Novartis Research Foundation, San Diego, California 92121, USA

†Department of Cell Biology, The Scripps Research Institute, La Jolla, California 92037, USA

e-mail: ardem@scripps.edu

- Caterina, M. J. & Julius, D. *Annu. Rev. Neurosci.* **24**, 487–517 (2001).
- Guler, A. D. et al. *J. Neurosci.* **22**, 6408–6414 (2002).
- Peier, A. M. et al. *Science* **296**, 2046–2049 (2002).
- McKemy, D. D., Neuhauser, W. M. & Julius, D. *Nature* **416**, 52–58 (2002).
- Peier, A. M. et al. *Cell* **108**, 705–715 (2002).
- Story, G. M. et al. *Cell* **112**, 819–829 (2003).
- Sayed, O. & Benzer, S. *Proc. Natl Acad. Sci. USA* **93**, 6079–6084 (1996).
- Liu, L., Yermolaieva, O., Johnson, W. A., Abboud, F. M. & Welsh, M. J. *Nature Neurosci.* **6**, 267–273 (2003).
- Tobin, D. et al. *Neuron* **35**, 307–318 (2002).
- Tracey, W. D., Wilson, R. I., Laurent, G. & Benzer, S. *Cell* **113**, 261–273 (2003).
- Walker, R. G., Willingham, A. T. & Zuker, C. S. *Science* **287**, 2229–2234 (2000).

Competing financial interests: declared none.

RNA interference

Producing decaffeinated coffee plants

The demand for decaffeinated coffee is increasing because the stimulatory effects of caffeine can adversely affect sensitive individuals by triggering palpitations, increased blood pressure and insomnia¹. Three *N*-methyltransferase enzymes are involved in caffeine biosynthesis in coffee plants — *CaXMT1*, *CaMXMT1* (theobromine synthase) and *CaDXMT1* (caffeine synthase), which successively add methyl groups to xanthosine in converting it into caffeine^{2–4}. Here we describe the construction of transgenic coffee plants in which expression of the gene encoding theobromine synthase (*CaXMT1*) is repressed by RNA interference (RNAi). The caffeine content of these plants is reduced by up to 70%, indicating that it should be feasible to produce coffee beans that are intrinsically deficient in caffeine.

Specific sequences in the 3' untranslated region (UTR) of *CaXMT1* messenger RNA were selected for construction of RNAi short and long fragments (Fig. 1). We transformed *Agrobacterium tumefaciens* EHA101 cells with these constructs and then used them to transform *Coffea canephora*⁵. After 2–4 months of culture, most infected tissues turned brown

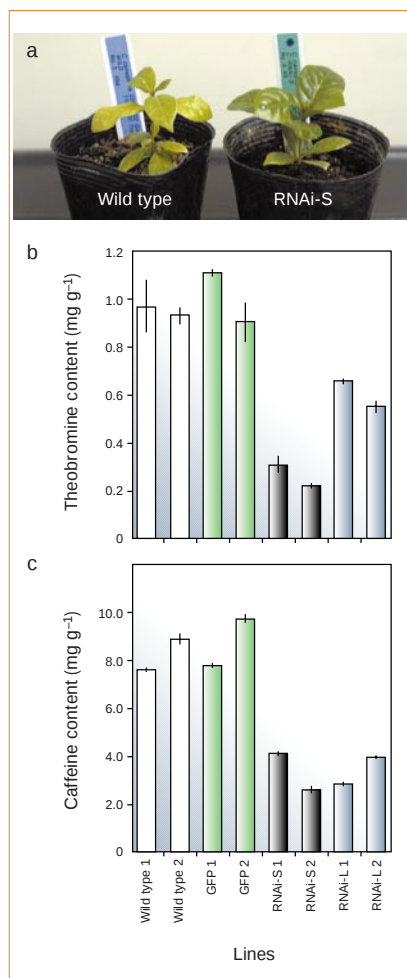


Figure 1 Properties of decaffeinated transgenic coffee leaves.

a, One-year-old somatic seedlings of *Coffea canephora* from wild-type (left) and RNAi-transgenic (right) plants. **b**, **c**, Endogenous theobromine and caffeine, respectively, in mg per g of fresh plant tissue, in different somatic seedlings of *C. canephora*, as detected by high-performance liquid chromatography². Mean values were calculated from six independent measurements per line. Short RNAi fragments (RNAi-S) were constructed using 139-base-pair (bp) corresponding to nucleotide positions 1,139–1,277 and 161-bp (positions 1,117–1,277) sequences of *CaXMT1* (GenBank accession number AB048794), with an intervening 517-bp β -glucuronidase (*GUS*) fragment as spacer; long RNAi fragments (RNAi-L) contained two identical sequences of 332 bp (positions 946–1,277) separated by a 517-bp *GUS* fragment. The resulting constructs were inserted into a pBIH1-IG vector⁶; the control construct contained a green fluorescent protein gene (*GFP*). Somatic embryos of *C. canephora* were grown on modified half-strength Murashige and Skoog medium containing 20 μ M 2-isopentenyladenine.

and necrotic; however, it was possible to regenerate hygromycin-resistant cells from these tissues. Seedlings were then cultured as described⁵.

More than 35 transgenic somatic seedlings were obtained from each transformant, each containing short or long RNAi fragments or a control gene encoding green fluorescent protein (*GFP*). The phenotypes were apparently normal when compared with the wild-type plant (Fig. 1a).

Young leaves of one-year-old seedlings were collected 2–3 weeks after flushing and their purine alkaloid content was measured. The wild-type and transgenic lines that expressed GFP contained similar amounts of endogenous theobromine and caffeine (about 1 and 8 mg per g of fresh plant tissue, respectively; Fig. 1b, c). By contrast, young leaves of transgenic lines expressing RNAi showed a 30–80% reduction in theobromine content (Fig. 1b) and a 50–70% reduction in caffeine content (Fig. 1c) in comparison with the controls.

At present, coffee is decaffeinated industrially, but the process is expensive and the flavour of the product is poor⁶ — problems that could potentially be overcome by the genetic engineering of coffee plants^{3,6}. As *CaMXMT1* is expressed in young leaves, buds and immature fruits⁴, the transgenic plants described here should yield coffee beans that are essentially normal apart from their low caffeine content at maturity.

We are now applying this RNAi-based technique to *C. arabica*, which produces high-quality Arabica coffee and accounts for roughly 70% of the world market. Our method not only shortens the breeding period, which is more than 25 years for conventional crossing, but also opens the way to develop new species of coffee plant.

Shinjiro Ogita, Hirota Uefuji, Yube Yamaguchi, Nozomu Koizumi, Hiroshi Sano
Research and Education Centre for Genetic Information, Nara Institute of Science and Technology, Nara 630-0192, Japan
e-mail: sano@gtc.aist-nara.ac.jp

- www.ico.org/frame5et/coffset.htm
- Ashihara, H., Monteiro, A. M., Gillies, F. M. & Crozier, A. *Plant Physiol.* **111**, 747–753 (1996).
- Ogawa, M., Herai, Y., Koizumi, K., Kusano, T. & Sano, H. *J. Biol. Chem.* **276**, 8213–8218 (2001).
- Uefuji, H., Ogita, S., Yamaguchi, Y., Koizumi, N. & Sano, H. *Plant Physiol.* **132**, 372–380 (2003).
- Hatanaka, T., Choi, Y. E., Kusano, T. & Sano, H. *Plant Cell Rep.* **19**, 106–110 (1999).
- Ashihara, H. & Crozier, A. *Trends Plant Sci.* **6**, 407–413 (2001).
- Ohta, S., Mita, S., Hattori, T. & Nakamura, K. *Plant Cell Physiol.* **31**, 805–813 (1990).

Competing financial interests: declared none.

COMMUNICATIONS ARISING

Ecology

Mycorrhizal weathering in base-poor forests

Minerals in soil contain many small pores, which has led to the suggestion that trees are able to ‘mine’ essential nutrients such as calcium through their association with symbiotic mycorrhizae, thereby bypassing the exchangeable calcium pool in the soil^{1,2}. On the basis of the calcium-to-strontium (Ca/Sr) ratios in the foliage of trees at Hubbard Brook (an experimental forest), Blum *et al.* suggest that these trees have direct access to calcium from apatite (calcium phosphate) in lower soil horizons

Table 1 Mean molar Ca/Sr ratios in tree parts

	Roots	Wood	Branches	Bark	Foliage
White pine	396 (148)	199 (78)	880 (432)	918 (130)	3,175 (1,858)
Eastern hemlock	814 (312)	275 (55)	756 (201)	831 (233)	2,655 (1,367)
White cedar	1,198 (661)	135 (8)	1,029 (200)	1,009 (421)	334 (123)
Black spruce	369 (151)	131 (27)	330 (90)	503 (154)	569 (189)
Balsam fir	511 (237)	188 (69)	402 (93)	500 (108)	802 (192)
Red maple	348 (128)	118 (22)	220 (31)	330 (59)	640 (273)
Red oak	349 (96)	126 (21)	800 (86)	629 (186)	1,626 (660)
Large tooth aspen	395 (255)	91 (18)	212 (17)	345 (91)	358 (64)
White birch	330 (81)	99 (16)	236 (61)	325 (50)	335 (108)

Data are from nine tree species at Plastic Lake, Ontario (45° 11', 78° 50'). Latin names for these species (top to bottom) are *Pinus strobus*, *Tsuga canadensis*, *Thuja occidentalis*, *Picea mariana*, *Abies balsamea*, *Acer rubrum*, *Quercus rubra*, *Populus grandidentata* and *Betula papyrifera*. Standard deviations are given in parentheses (*n*, 5–10).

through their mycorrhizal associations³. But because Ca/Sr ratios can vary greatly in different parts of a plant, Ca/Sr ratios in foliage are poor indicators of the source of calcium for trees. Differences in Ca/Sr ratios in tree foliage should be interpreted with caution until more is known about the cycling of these elements in forests.

Depletion of the exchangeable calcium pool in soil is of widespread concern and may have serious implications for future forest health and productivity^{4–8}. Blum *et al.* propose that apatite-derived calcium is used largely by ectomycorrhizal tree species, and that mycorrhizae may weather apatite and absorb the released ions directly, without ions entering the exchangeable pool³.

The mineralogy of the shallow soils and bedrock at Plastic Lake (referred to here as PC1) in south-central Ontario is dominated by quartz, plagioclase, K-feldspar, hornblende, vermiculite, imogolite and goethite; there is no evidence that apatite is present at PC1 (ref. 9). Furthermore, base-cation mass-balance budgets that consider only silicate weathering are consistent with the measured changes in stream water and soil chemistry, which together indicate that substantial calcium losses have occurred from the exchangeable pool over the past two decades¹⁰.

Despite the absence of calcium-rich minerals at PC1, molar Ca/Sr ratios in tree foliage at PC1 vary widely and range from about 330 to 3,200 (Table 1), which is similar to the range (about 500 to 2,200) reported at Hubbard Brook³. As shown previously¹¹, we found that Ca/Sr ratios vary widely across plant parts within a single tree species (Table 1).

The annual cycling of calcium in the above-ground biomass (litter fall plus foliar leaching) of temperate forests in eastern North America is typically an order of magnitude greater than the rate of calcium weathering expected in base-poor forests^{12,13}. Because the 'mining' of calcium cannot supply a large proportion of a forest's annual calcium demand, the direct uptake of calcium from minerals will not protect trees from the negative effects of low calcium or high aluminium availability that may occur in acidified soils^{1,3}. The ecological significance of direct calcium uptake

through symbiotic mycorrhizal association lies in whether this process increases the annual calcium weathering rate, which will determine the size of the exchangeable calcium pool under given acid-deposition and harvesting conditions.

Shaun A. Watmough, Peter J. Dillon

Environmental and Resource Studies Department, Trent University, Peterborough, Ontario K9J 7B8, Canada

e-mail: swatmough@trentu.ca

1. Jongmans, A. G. *et al.* *Nature* **389**, 682–683 (1997).
2. Van Breemen, N. *et al.* *Biogeochemistry* **49**, 53–67 (2000).
3. Blum, J. D. *et al.* *Nature* **417**, 729–731 (2002).
4. Likens, G. E., Driscoll, C. T. & Buso, D. C. *Science* **272**, 244–246 (1996).
5. Johnson, C. E., Driscoll, C. T., Siccama, T. G. & Likens, G. E. *Ecosystems* **3**, 159–184 (2000).
6. Federer, C. A. *et al.* *Environ. Mgmt* **13**, 593–601 (1989).
7. Sverdrup, H. & Rosen, K. *For. Ecol. Mgmt* **110**, 221–236 (1998).
8. Lawrence, G. B., David, M. B., Bailey, S. W. & Shortle, W. C. *Biogeochemistry* **38**, 19–39 (1997).
9. Kirkwood, D. E. & Nesbitt, H. W. *Geochim. Cosmochim. Acta* **55**, 1295–1308 (1991).
10. Watmough, S. A. & Dillon, P. J. *For. Ecol. Mgmt* **177**, 155–177 (2003).
11. Poszwa, A., Dambrine, E., Pollier, B. & Atteia, O. *Plant Soil* **225**, 299–310 (2000).
12. Likens, G. E. *et al.* *Biogeochemistry* **41**, 89–173 (1998).
13. Duchesne, L., Ouimet, R., Camiré, C. & Houle, D. *Can. J. For. Res.* **31**, 333–334 (2001).

Blum *et al.* reply — We identified apatite as an important reservoir of calcium in the soil horizons termed Bs and C at Hubbard Brook experimental forest (HBEF) and suggested that it could exceed the size of the soil-exchange pool^{1,2}. Apatite has high calcium-to-strontium (Ca/Sr) and phosphorus-to-calcium ratios, distinct from silicate mineral and atmospheric sources of calcium. The higher Ca/Sr ratios in foliage of ectomycorrhizal tree species (compared with non-ectomycorrhizal species) implied that calcium in apatite might be more accessible to ectomycorrhizal species¹. Some preferential plant uptake of calcium over strontium was expected^{1,3}, so we reported the proportions of apatite-derived foliar calcium as maximum values¹. Previous work indicated that species differences in Ca/Sr were more directly related to soil mineralogy (silicate versus calcite) than to preferential uptake that depended on species^{3,4}.

Ca/Sr ratios in HBEF roots and wood also reveal higher Ca/Sr in ectomycorrhizal

trees (J.D.B. *et al.*, unpublished results), whereas Watmough and Dillon do not find consistent variations in the Ca/Sr ratio of various tree parts from different species at Plastic Lake, Ontario. Further work is needed to understand these intra-species variations, and we agree that Ca/Sr ratios in foliage should be interpreted with caution, particularly where detailed soil chemical and mineralogical data are not available.

Watmough and Dillon show that pine, hemlock and red oak (all ectomycorrhizal species) at Plastic Lake have high foliar Ca/Sr ratios. Although apatite has not been identified as a major soil mineral at Plastic Lake⁵, virtually all crystalline silicate rocks contain small amounts of apatite⁶ and/or calcite⁷, including those throughout Ontario⁶. Soil phosphorus concentrations and/or cathodoluminescence images have not been reported from Plastic Lake and would be necessary to identify minute inclusions of apatite and/or calcite, which are easily weathered trace minerals with high Ca/Sr ratios.

Our results show that apatite accounts for about 20% of the calcium in the HBEF soil parent material¹, and more than 20% of the calcium released by weathering since deglaciation was derived from apatite. Nevertheless, most of the annual calcium uptake by trees is recycled through detritus, with only a small annual contribution of newly weathered calcium. Therefore, although the apatite calcium pool in HBEF soils changes our view of 'plant-available' calcium, direct uptake of calcium from minerals will not fully protect trees from the negative effects of calcium depletion in acidified soils. We agree that the ecological significance of calcium release by ectomycorrhizal-driven weathering lies in its effect on calcium-weathering rates — indeed, organic acids of low relative molecular mass (which are released by ectomycorrhizal hyphae⁸) accelerate apatite weathering markedly compared with inorganic acids⁹.

Joel D. Blum*, Andrea Klau, Carmen A. Nezat, Charles T. Driscoll, Chris E. Johnson, Thomas G. Siccama, Christopher Eagar, Timothy J. Fahey, Gene E. Likens

**Department of Geological Sciences, University of Michigan, Ann Arbor, Michigan 48109, USA
e-mail: jdblum@umich.edu*

1. Blum, J. D. *et al.* *Nature* **417**, 729–731 (2002).
2. Hamburg, S. P., Yanai, R. D., Arthur, M. A., Blum, J. D. & Siccama, T. G. *Ecosystems* (in the press).
3. Blum, J. D., Taliaferro, H., Weiss, M. T. & Holmes, R. T. *Biogeochemistry* **49**, 87–101 (2000).
4. Poszwa, A., Dambrine, E., Pollier, B. & Atteia, O. *Plant Soil* **225**, 299–310 (2000).
5. Kirkwood, D. E. & Nesbitt, H. W. *Geochim. Cosmochim. Acta* **55**, 1295–1308 (1991).
6. Shaw, D. M., Reilly, G. A., Muiysson, J. R., Pattenden, G. E. & Cambell, F. E. *Can. J. Earth Sci.* **4**, 829–853 (1967).
7. White, A. F., Bullen, T. D., Vivit, D. V., Schulz, M. S. & Clow, D. W. *Geochim. Cosmochim. Acta* **63**, 1939–1953 (1999).
8. Van Breeman, N. *et al.* *Biogeochemistry* **49**, 53–67 (2000).
9. Welch, S. A., Taunton, A. E. & Banfield, J. F. *Geomicrobiol. J.* **19**, 343–367 (2002).

The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes

Helen Skaletsky*, Tomoko Kuroda-Kawaguchi*, Patrick J. Minx†, Holland S. Cordum†, LaDeana Hillier†, Laura G. Brown*, Sjoerd Repping‡, Tatyana Pyntikova*, Johar Ali†, Tamberlyn Bieri†, Asif Chinwalla†, Andrew Delehaunty†, Kim Delehaunty†, Hui Du†, Ginger Fewell†, Lucinda Fulton†, Robert Fulton†, Tina Graves†, Shun-Fang Hou†, Philip Latrielle†, Shawn Leonard†, Elaine Mardis†, Rachel Maupin†, John McPherson†, Tracie Miner†, William Nash†, Christine Nguyen†, Philip Ozersky†, Kymberlie Pepin†, Susan Rock†, Tracy Rohlfing†, Kelsi Scott†, Brian Schultz†, Cindy Strong†, Aye Tin-Wollam†, Shiaw-Pyng Yang†, Robert H. Waterston†, Richard K. Wilson†, Steve Rozen* & David C. Page*

* Howard Hughes Medical Institute, Whitehead Institute, and Department of Biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA

† Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Boulevard, St Louis, Missouri 63108, USA

‡ Center for Reproductive Medicine, Department of Gynaecology and Obstetrics, Academic Medical Centre, Amsterdam 1105 AZ, the Netherlands

The male-specific region of the Y chromosome, the MSY, differentiates the sexes and comprises 95% of the chromosome's length. Here, we report that the MSY is a mosaic of heterochromatic sequences and three classes of euchromatic sequences: X-transposed, X-degenerate and ampliconic. These classes contain all 156 known transcription units, which include 78 protein-coding genes that collectively encode 27 distinct proteins. The X-transposed sequences exhibit 99% identity to the X chromosome. The X-degenerate sequences are remnants of ancient autosomes from which the modern X and Y chromosomes evolved. The ampliconic class includes large regions (about 30% of the MSY euchromatin) where sequence pairs show greater than 99.9% identity, which is maintained by frequent gene conversion (non-reciprocal transfer). The most prominent features here are eight massive palindromes, at least six of which contain testis genes.

The history of human Y chromosome research can be divided into three eras. The first era focused on mendelian examination of human family trees. In the opening decades of the twentieth century, proponents of Mendel's concept of the gene observed three modes of inheritance in our species: autosomal recessive, autosomal dominant and X-linked recessive. Contemporaneously, other scholars sought to identify traits that exhibited Y-linked (father to son) transmission. These scholars erroneously claimed success, presenting family trees purported to demonstrate that Y-chromosomal genes were responsible for hairy ears, scaly skin and other traits. Meanwhile, light microscopic studies of human cells provided strong physical evidence of the existence of a male-specific chromosome¹. By 1950, studies of human pedigrees reported at least 17 Y-linked traits².

The second era was dominated by the view that the Y chromosome was a genetic wasteland, based on the debunking of earlier studies and a dearth of new evidence for genes. In the 1950s, Stern systematically exposed critical flaws in each of the preceding pedigree studies and dismissed them². In 1959, Jacobs' study of Klinefelter (XXY) males³ and Ford's research on Turner (XO) females⁴ demonstrated that the Y chromosome carries a pivotal sex-determining gene, but this gene was considered to be an exception on a generally desolate chromosome. In the 1960s, Ohno proposed that the mammalian X and Y chromosomes had evolved from an ordinary pair of autosomes⁵. Ohno speculated that the X chromosome had retained the ancestral autosome's gene content whereas the Y chromosome had lost all but perhaps one gene involved in sex determination. Thus emerged the understanding of the human Y chromosome as a profoundly degenerate X chromosome.

The hallmark of the third and present era has been the application of recombinant DNA and genomic technologies to the Y chromo-

some, culminating in molecularly based conclusions about its genes. In recent decades, an understanding of the Y chromosome's biological functions has begun to emerge from DNA studies of individuals with partial Y chromosomes, coupled with molecular characterization of Y-linked genes implicated in gonadal sex reversal, Turner syndrome, graft rejection and spermatogenic failure⁶. Genomic studies revealed that the Y chromosome contains a region, comprising 95% of its length, where there is no X–Y crossing over. This region came to be known as the non-recombining region, or NRY, although our discovery of abundant recombination, as reported here and in the accompanying manuscript, compels us to rename it the male-specific region, or MSY⁷. The MSY is flanked on both sides by pseudoautosomal regions, where X–Y crossing over is a normal and frequent event in male meiosis (see Supplementary Note 1).

Previous efforts to construct accurate, high-resolution physical maps of the MSY had been stymied by an abundance of lengthy, intrachromosomal repetitive sequences, or amplicons⁸. To overcome this difficulty, we identified minute variations between amplicon copies, and then highlighted these minute variants (sequence family variants⁹) as markers to be ordered with respect to one another, yielding a map amenable to iterative refinement. However, the minute variants could only be found by fully and accurately sequencing and comparing near-identical amplicon copies. Thus, in our effort to determine the nucleotide sequence of the MSY, mapping and sequencing activities were fused into a single, iterative analytic process. We have previously reported the physical map that emerged from these efforts¹⁰. Here we report the sequencing of the MSY.

We mapped and sequenced a tiling path of 220 bacterial artificial chromosome (BAC) clones, each containing a portion of the MSY from the same individual. We used only one man's Y chromosome

to prevent any allelic variation, or polymorphism, from confounding our search for minute sequence variation between amplicon copies. (MSY amplicon copies can differ as little in sequence as two Y chromosomes chosen at random from the population¹¹.) We chose to sequence highly redundant BACs, especially in amplicon-rich regions: about 12.7 million (roughly 60%) of the euchromatic nucleotides were sequenced in at least two independent BAC clones. This redundancy allowed us to refine and validate the MSY sequence by exhaustively investigating, and in most cases resolving, sequence discrepancies between overlapping BACs.

Sequencing of euchromatic and heterochromatic regions

We begin with a statistical synopsis of the MSY sequence, considering the euchromatic and heterochromatic portions separately. (In this analysis, we have equated satellite sequences with heterochromatin, and all other sequences with euchromatin.) The product of our present research is a 'reference' sequence from one man's Y chromosome. A full description of the nature and extent of

Y chromosome variation in human populations must await future studies. We and our colleagues have previously reported the nucleotide sequence of two portions of the MSY (the *AZFa* and *AZFc* regions^{12,13}). We have incorporated this previously reported sequence data in our present analysis of the entire MSY.

The MSY's euchromatic DNA sequences total roughly 23 megabases (Mb), including 8 Mb on the short arm (Yp) and 14.5 Mb on the long arm (Yq) (Fig. 1). We obtained finished sequence, with an estimated error rate of about 1 per 10⁵ nucleotides, for all MSY euchromatin, with two known exceptions. First, there remain two gaps, each of which is roughly 50 kilobases (kb) long as judged by chromosomal fluorescence *in situ* hybridization (FISH) (Supplementary Fig. 1). Second, we obtained representative but incomplete sequence for a tandem array that spans roughly 0.7 Mb on Yp. We estimate that we obtained finished nucleotide sequence for roughly 97% of the MSY euchromatin, and that we captured 99% of the sequence complexity of MSY's euchromatin.

So far, efforts to gain sequence-based understanding of human chromosomes have largely by-passed heterochromatic regions (refs 14, 15; see also Supplementary Note 2), including a large block of heterochromatic sequences found in the centromeric region of every nuclear chromosome¹⁶. In addition to its centromeric heterochromatin (approximately 1 Mb, ref. 17), the Y chromosome was previously shown to contain a second, much longer heterochromatic block (roughly 40 Mb) that comprises the bulk of the distal long arm (Fig. 1; see also Supplementary Note 3). In the course of the present sequencing project, we discovered and characterized a third heterochromatic block—a sharply demarcated island that spans approximately 400 kb, comprises >3,000 tandem repeats of 125 base pairs (bp), and interrupts the euchromatic sequences of proximal Yq (Figs 1 and 2). The other two heterochromatic blocks also consist of massively amplified tandem repeats of low sequence complexity. We attempted to sequence BACs spanning the boundaries and representing the body of each of the three heterochromatic blocks. We succeeded, with the exception that the distal boundary of the major heterochromatic region, on distal Yq, was not identified with certainty (Supplementary Fig. 1). In total, we found that the heterochromatin of MSY encompasses at least six distinct sequence species (Table 1), each of which form long, homogeneous tandem arrays. Our findings are detailed in Supplementary Note 4 and Supplementary Fig. 3.

A catalogue of genes and transcription units

With a comprehensive reference sequence of the MSY in hand, we set out to catalogue systematically the genes of the MSY. We electronically identified and manually examined all matches to previously reported MSY genes. Furthermore, we used polymerase chain reaction with reverse transcription (RT-PCR) and/or sequencing of complementary DNA clones to evaluate electronic matches to publicly available expressed sequence tags (ESTs), as well as potential genes that were predicted using GenomeScan software¹⁸. For all experimentally verified genes whose expression patterns had not been reported previously, we tested for expression in diverse human tissues by RT-PCR and subsequent sequencing of RT-PCR products.

We found that the MSY includes at least 156 transcription units, half of which probably encode proteins (Table 2 and Figs 2, 3; see also Supplementary Tables 1 and 2). All 156 transcription units identified are located in euchromatic sequences. We have no evidence of transcription of the MSY heterochromatin. Of the approximately 78 protein-coding units, about 60 are members of nine different MSY-specific gene families, each characterized by >98% nucleotide identity among family members, in both exons and introns. The remaining 18 protein-coding genes are present in one copy each in the MSY. (These include two genes, *RPS4Y1* and *RPS4Y2*, that exhibit 93.6% nucleotide identity in coding exons but are much more diverged in introns.) Thus, the MSY seems to

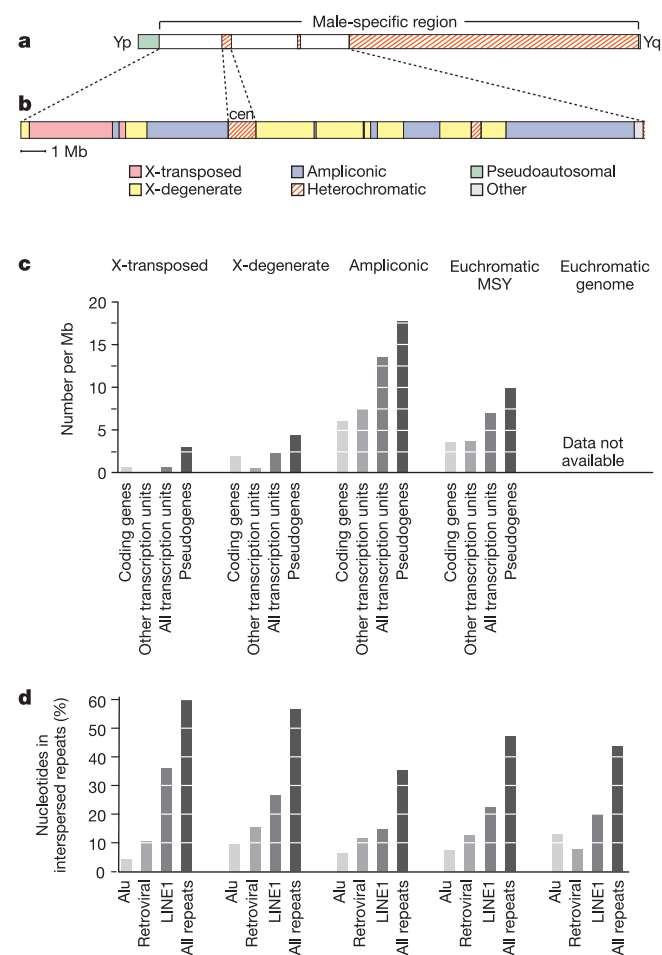


Figure 1 The male-specific region of the Y chromosome. **a**, Schematic representation of the whole chromosome, including the pseudoautosomal and heterochromatic regions. **b**, Enlarged view of a 24-Mb portion of the MSY, extending from the proximal boundary of the Yp pseudoautosomal region to the proximal boundary of the large heterochromatic region of Yq. Shown are three classes of euchromatic sequences, as well as heterochromatic sequences. A 1-Mb bar indicates the scale of the diagram. **c**, **d**, Gene, pseudogene and interspersed repeat content of three euchromatic sequence classes. **c**, Densities (numbers per Mb) of coding genes, non-coding transcription units, total transcription units and pseudogenes. **d**, Percentages of nucleotides contained in Alu, retroviral, LINE1 and total interspersed repeats. The data shown in **c** and **d** are available in numerical form in Supplementary Tables 6 and 7. Supplementary Table 6 also provides information about the size and (G + C) content of each sequence class.

encode at least 27 distinct proteins or protein families.

Furthermore, the MSY includes at least 78 transcription units for which strong evidence of protein coding is lacking; many of these transcription units are probably non-coding. Of these 78 transcription units, 13 occur in single copy in the MSY and the remaining 65 are members of 15 MSY-specific families. Considering together both coding and non-coding transcription units, the MSY appears to contain 24 MSY-specific families, which collectively account for 125 of the 156 MSY transcription units identified so far.

On the basis of earlier experiments, most of the genes of the MSY were thought to fall into two functional classes, with genes in the first group expressed throughout the body, in many organs, and genes in the second group expressed predominantly or exclusively in testes¹⁹. Our present catalogue of MSY genes and their patterns of tissue expression (Table 2) corroborate this model. Of the MSY's 27 distinct protein-coding genes or gene families identified so far, 12 are expressed ubiquitously and 11 are expressed exclusively or predominantly in testes.

Three classes of sequences in the MSY euchromatin

We find that nearly all of the euchromatic sequences fall into three classes, which we have named X-transposed, X-degenerate and ampliconic. As shown in Figs 1 and 2, the MSY euchromatin is a patchwork of these three sequence classes. The characteristics of the classes are summarized in Fig. 4.

The X-transposed sequences are 99% identical to DNA sequences in Xq21, a band in the midst of the long arm of the human X chromosome. The X-transposed sequences are so named because their presence in the human MSY is the result of a massive X-to-Y transposition that occurred about 3–4 million years ago, after the divergence of the human and chimpanzee lineages^{20–22}. Subsequently, an inversion within the MSY short arm cleaved the X-transposed block into two non-contiguous segments, as observed in the modern MSY (Figs 1 and 2)^{21,22}. The X-transposed sequences do not participate in X–Y crossing over during male meiosis, distinguishing them from the pseudoautosomal sequences found in the telomeric regions of the human X and Y chromosomes.

Within the X-transposed segments, which have a combined length of 3.4 Mb, we identified only two genes, both of which have homologues in Xq21 (Table 2). Thus the X-transposed sequences exhibit the lowest density of genes among the three sequence classes in the MSY euchromatin (Figs 1 and 3), as well as the highest density of interspersed repeat elements (Fig. 1). In particular, long interspersed nuclear element 1 (LINE1) elements account for 36% of all X-transposed sequence, or nearly twice the genome average of 20%^{14,15}. As expected, low gene density and high repeat density also characterize the homologous sequence block in Xq21.

In contrast to the X-transposed sequence blocks, the X-degenerate segments of the MSY are dotted with single-copy gene or pseudogene homologues of 27 different X-linked genes. These single-copy MSY genes and pseudogenes display between 60%

and 96% nucleotide sequence identity to their X-linked homologues, and they seem to be surviving relics of ancient autosomes from which the X and Y chromosomes co-evolved, as explained below. In 13 cases, the MSY homologue is a pseudogene with sequence similarity to exons and introns of the functional X homologue (Supplementary Table 3). In the remaining 14 cases, the MSY homologue seems to be a transcribed, functional gene, and the X- and Y-linked genes encode very similar but non-identical protein isoforms (Table 2 and Figs 2, 3). These include two cases in which a functional X-linked gene has two expressed homologues in the MSY. The Y-linked genes *RPS4Y1* and *RPS4Y2* are full-length homologues of the X-linked gene *RPS4X*, and they apparently encode two different, full-length isoforms of ribosomal protein S4. In contrast, the Y-linked genes *CYorf15A* and *CYorf15B* are homologous to, respectively, 5' and 3' portions of the X-linked gene *CXorf15*, and they apparently encode proteins homologous to, respectively, amino- and carboxy-terminal portions of the predicted CXORF15 protein (Supplementary Fig. 4). Together, the X-degenerate sequences encode 16 of the MSY's 27 distinct proteins or protein families.

Notably, all 12 ubiquitously expressed MSY genes reside in the X-degenerate regions; no such genes have been identified elsewhere in the MSY. Conversely, among the 11 MSY genes found to be expressed predominantly in testes, only one gene, the sex-determining *SRY*, is X-degenerate.

The third class of euchromatic sequences, the ampliconic segments, are composed largely of sequences that exhibit marked similarity—as much as 99.9% identity over tens or hundreds of kilobases—to other sequences in the MSY. We refer to these long, MSY-specific repeat units, of which there are many families, as amplicons. The amplicons are located in seven segments that are scattered across the euchromatic long arm and proximal short arm (Figs 1 and 2), and whose combined length is 10.2 Mb.

We identified these ampliconic regions through a comprehensive analysis of similarities within the sequenced portions of the MSY. We calculated the percentage nucleotide identity between all pairs of known MSY sequences and then plotted the data in two ways. First we determined, at each point along the length of the sequenced MSY, the highest intrachromosomal similarity. The resulting graph (Fig. 5c) identifies the ampliconic regions as those where intrachromosomal identity, over stretches of 50 kb or more, generally exceeds 50%. Notably, 60% (6.1 Mb) of the ampliconic sequences exhibit intrachromosomal identities of 99.9% or greater.

A more spatially detailed representation of intrachromosomal similarities is shown in Fig. 5a, which records the locations of all MSY sequence pairs characterized by at least 65% identity within a sliding window of 2,000 nucleotides. After heterochromatic and LINE1 repeats have been accounted for, the MSY is seen to contain many long stretches of sequence that are similar to those elsewhere in the MSY. As shown in the inset to Fig. 5a, the triangular plot can be broken down into two smaller triangles—one representing sequence comparisons within Yp, the other depicting comparisons

Table 1 Six sequence species in three MSY heterochromatic regions

Region	Locus symbol	Unit repeat	
		Primary	Secondary
Centromeric region	<i>DYZ3</i> <i>DYZ17*</i>	171 bp alphoid GGAAT	5,941 bp None observed
Proximal Yq (Yq11.22)	<i>DYZ19*</i>	125 bp	None observed
Distal Yq (Yq12)	<i>DYZ18*</i> <i>DYZ1</i> <i>DYZ2</i>	GGAAT GGAAT Insufficient sequence available; (A + T) rich	2,864 bp 3,584 bp ~2,470 bp†

See Supplementary Table 10 for a more detailed version of this Table, incorporating references and the sequences of repeat units.

*First reported in this study.

†On the basis of restriction endonuclease digestion, not sequence analysis.

within Yq—and a rectangle depicting comparisons between Yp and Yq. Scrutiny of these Yp, Yq and Yp–Yq components of the plot reveals a wealth of sequence similarities within and between ampliconic segments on both arms of the chromosome.

The ampliconic sequences exhibit by far the highest density of genes, both coding and non-coding, among the three sequence classes in the MSY euchromatin (Figs 1 and 3). We identified nine distinct MSY-specific protein-coding gene families, with copy numbers ranging from two (*VCY*, *XKRY*, *HSFY*, *PRY*) to three (*BPY2*) to four (*CDY*, *DAZ*) to six (*RBMV*) to approximately 35 (*TSPY*) (Table 2 and Figs 2, 3). (These copy numbers pertain to the particular Y chromosome that we sequenced; they may vary in human populations.) In aggregate, these nine coding families encompass roughly 60 transcription units. Furthermore, the ampliconic sequences include at least 75 other transcription units for which strong evidence of protein coding is lacking (Figs 2 and 3; see also Supplementary Table 2). Of these 75 putative non-coding transcription units, 65 are members of 15 MSY-specific families, and the remaining 10 occur in single copy. Considering together both coding and non-coding elements, the ampliconic sequences contain 135 of the 156 MSY transcription units identified so far.

In contrast to the ubiquitous expression of most X-degenerate genes, the ampliconic genes and transcription units show highly restricted expression (Table 2). All nine protein-coding families in the ampliconic regions are expressed predominantly or exclusively in testes, as are most of the regions' non-coding transcription units.

Among the three euchromatic sequence classes, the ampliconic sequences exhibit by far the lowest densities of LINE1 and total interspersed repeat elements (Fig. 1). Indeed, the interspersed repeat content of the MSY's ampliconic sequences (35%) is far below the mean for the human genome (44%; z -test yields $P \ll 0.000001$).

Eight palindromes comprising 25% of MSY euchromatin

The most pronounced structural features of the ampliconic regions of Yq are eight massive palindromes (Table 3). In the dot plot of Fig. 5a, the longer palindromes are visible as vertical blue lines that approach the baseline. An MSY map highlighting all eight palindromes is shown in Fig. 3a. In all eight palindromes, the arms are highly symmetrical, with arm-to-arm nucleotide identities of 99.94–99.997%. (By convention, these percentage identities refer only to nucleotide substitutions and do not take account of insertions and deletions by which palindrome arms differ.) The palindromes are long, their arms ranging from 9 kb to 1.45 Mb in length. They are imperfect in that each contains a unique, non-duplicated spacer, 2–170 kb in length, at its centre. Palindrome P1 is particularly spectacular, having a span of 2.9 Mb, an arm-to-arm identity of 99.97%, and bearing two secondary palindromes (P1.1 and P1.2, each with a span of 24 kb) within its arms¹³. The eight palindromes collectively comprise 5.7 Mb, or one-quarter of the MSY euchromatin.

Six of the eight palindromes carry recognized protein-coding genes, all of which seem to be expressed specifically in testes (Fig. 3b). In all known cases of genes on MSY palindromes, identical or nearly identical gene copies exist on opposite arms of the palindrome. Of the nine multi-copy, protein-coding gene families identified so far in the MSY, eight have members on palindromes. Indeed, six families are located exclusively in palindromes. These include the *DAZ* genes, which exist in four copies—two in palindrome P1 and two in P2—and the *CDY* genes, which also occur in four copies—two in P1 and two in P5 (Fig. 3b). In addition, the palindromes contain at least seven families of apparently non-coding transcription units, all expressed exclusively or predominantly in testes (Fig. 3c).

In addition to the eight palindromes, the ampliconic regions of Yq and Yp contain five sets of more widely spaced inverted repeats with repeat lengths of 62–298 kb (Fig. 2; see also Supplementary Table 4). Three of these inverted repeat pairs (IR1, IR2 and IR3)

exhibit nucleotide identities of 99.66–99.95%. Inversion of the IR3 repeats, both located on Yp, was probably a direct consequence of the molecular evolutionary event that cleaved the X-transposed sequences into two non-contiguous segments (Supplementary Fig. 5). Subsequent homologous recombination between inverted IR3 repeats was responsible, we suspect, for a 3.6-Mb inversion polymorphism observed on the short arm of the modern Y chromosome (Supplementary Figs 5 and 6)¹⁰.

Transcriptionally active tandem arrays

In addition to palindromes and inverted repeats, the ampliconic regions of Yq and Yp contain a variety of long tandem arrays. Prominent among these are the newly identified NORF (no long open reading frame) clusters, which in aggregate account for about 622 kb on Yp and Yq, and the previously reported TSPY clusters, which comprise about 700 kb of Yp (Fig. 2). Triangular dot plots that highlight the regularities and relatively crisp borders of the NORF and TSPY arrays are shown in Supplementary Fig. 7 (see also Supplementary Note 5).

The NORF arrays are based on a repeat unit of 2.48 kb. A consensus sequence for the repeat is readily identifiable (Supplementary File 2), but the sequence of individual repeat elements typically diverges from that consensus by 14–20%. The NORF arrays are so named because they harbour a great diversity of spliced but apparently non-coding transcription units, including the *TTY1*, *TTY2*, *TTY6*, *TTY7*, *TTY8*, *TTY18*, *TTY19*, *TTY21* and *TTY22* families. Both strands of the NORF arrays are transcribed; 3' portions of the *TTY1* and *TTY2* transcripts are complementary (Supplementary Fig. 8).

The TSPY arrays are based on a 20.4-kb repeat unit²³ that encodes, on one strand, a previously identified protein, TSPY. A newly identified transcription unit, *CYorf16*, is found on the opposite

Figure 2 Sequence-based map of the MSY; a detailed view of the 24-Mb region shown in Fig. 1b. Background colours indicate the three classes of MSY euchromatic sequences: X-transposed (pink), X-degenerate (yellow) and ampliconic (blue), as well as heterochromatic (red stripes) and pseudoautosomal (green) sequences and NORF arrays (grey stripes). Two gaps in the sequence are indicated at the top edge of the diagram. **a**, Eight primary palindromes (P1–P8) and two secondary palindromes (P1.1 and P1.2). Diverging black arrows mark the left and right arms of each palindrome. Gaps between diverging arrows represent non-palindromic spacers at the centres of these structures. **b**, Near-perfect inverted repeats (non-palindromic), three in all (IR1 to IR3; Supplementary Table 4). In each case, the left and right arms exhibit >99.5% nucleotide identity. **c**, Other inverted repeats (non-palindromic). Grey arrows (IR4) denote two regions of >93% identity, one on Yp and one on Yq. Yellow arrows (IR5) denote four regions of >92% identity, all on Yq. **d**, Deletions of any of the four indicated regions—*AZF*a, *P5/proximal P1* (*AZF*b), *AZF*c, or *P5/distal P1*—cause spermatogenic failure^{13,43–46}. **e**, Previously reported genes and new, experimentally verified transcription units for which cDNA sequencing suggests protein-coding potential (Table 2). Plus (+) and minus (–) strand are indicated by the top or bottom row, respectively. **f**, See Fig. 5c. **g**, Scale (Mb). **h**, Sequences whose transcription has been verified (in this or previous studies) but for which there is little or no evidence of protein-coding potential (Supplementary Table 2). **i**, Previously reported pseudogenes and new, apparently non-transcribed homologues of known coding genes (Supplementary Table 3). **j**, (G + C) content (%) calculated in a 100-kb sliding window with 1-kb steps. **k**, Alu, LINE1 and human endogenous retroviral (HERV) repeat content, expressed as percentage of nucleotides, calculated in a 200-kb sliding window with 1-kb steps. **l**, 220 BAC clones completely or partially sequenced. Each bar represents size and position of one BAC clone, identified by the numeric portion of its GenBank accession number (in each case beginning with the prefix AC). Black bars represent finished sequences deposited in GenBank, where finished sequences are trimmed to retain only 200 bp of overlap with adjoining BACs. Grey bars represent the 'trimmings' of those BACs, not deposited in GenBank. Striped bars represent BACs whose sequence has not been finished but has been deposited in GenBank. See Supplementary Fig. 2 for a more detailed version of this figure. The composite sequence of the 24-Mb region studied is available as Supplementary File 1.

strand; its protein coding potential remains to be tested. Approximately 35 copies of this repeat unit—and hence 35 *TSPY* genes and 35 *CYorf16* transcription units—are found in a single, highly regular tandem array in proximal Yp (Fig. 2 and Supplementary Fig. 7d, e); here the sequences of individual repeat units rarely differ from the consensus by more than 1%. Furthermore, a single, isolated TSPY repeat unit, whose sequence diverges 3% from the consensus, is located more distally in Yp, embedded in the distal IR3 inverted repeat (Fig. 2). The 35-unit TSPY cluster is the largest and most homogeneous protein-coding tandem array identified so far in the human genome.

The evolution of the MSY

On the basis of our present findings and previous studies, we propose a model of MSY evolution that addresses all three euchromatic sequence classes (Figs 6 and 7). In developing the model, we will offer an evolutionary map of the MSY (Fig. 8). We will then consider the two largest and most gene-rich sequence classes—X-degenerate and ampliconic—arguing that two opposed evolutionary dynamics have been at work: gene decay versus gene acquisition and conservation. Throughout, we will propose decisive roles for

modulation of DNA recombination, both crossing over and gene conversion, in the evolution and on-going maintenance of the MSY (Fig. 9).

The human X and Y chromosomes are thought to have evolved from an ordinary pair of autosomes^{5,24}. Support for this hypothesis, and a proposed 300-million-year timeline for human sex chromosome evolution, have emerged from studies of modern X–Y gene pairs. In this context, investigators have interpreted the X–Y gene pairs as surviving ‘fossils’ where extensive sequence identity between ancestral X and Y chromosomes once existed^{25,26}. Our present sequencing of the MSY euchromatin expands the catalogue of known X–Y gene pair fossils, providing opportunity to re-examine models developed in earlier studies.

Evolutionary stratification of X–Y genes

Lahn and Page previously studied the evolutionary ages of X–Y gene pairs, as measured by synonymous X–Y nucleotide divergence, or K_s (ref. 26). They reasoned that X–Y differentiation would have begun only after X–Y crossing over ceased. They observed a strong correlation between the age (K_s) of individual X–Y gene pairs and the locations of their X members on the human X chromosome.

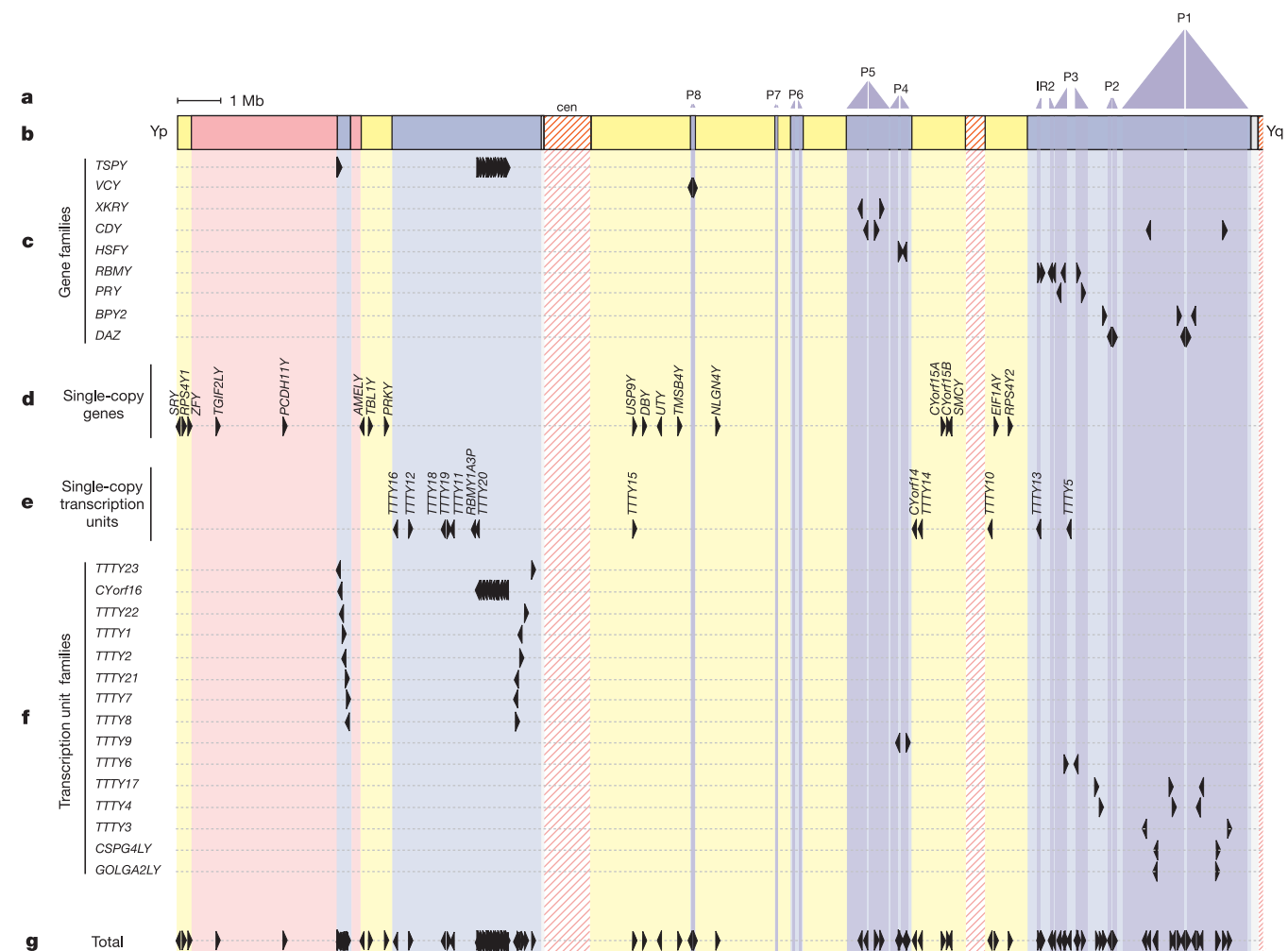


Figure 3 MSY genes, transcription units and palindromes. **a**, Triangles denote sizes and locations of arms of eight palindromes (P1–P8) and of IR2 inverted repeats (whose arms exhibit 99.95% identity). Gaps between opposed triangles represent the non-duplicated ‘spacers’ between palindrome arms. **b**, MSY schematic, as in Fig. 1b. **c**, Nine families of protein-coding genes. Solid triangles denote apparently intact genes (5′ to 3′ polarity indicated); pseudogenes are not shown. **d**, Single-copy protein-coding genes. **e**, Single-copy transcription units. These give rise to spliced but apparently non-coding transcripts. **f**, Fifteen families of transcription units. **g**, Merged map of all genes and transcription units.

indicated); pseudogenes are not shown. **d**, Single-copy protein-coding genes. **e**, Single-copy transcription units. These give rise to spliced but apparently non-coding transcripts. **f**, Fifteen families of transcription units. **g**, Merged map of all genes and transcription units.

Among the 19 X–Y gene pairs studied, age increased in a stepwise fashion along the length of the X chromosome, in four ‘evolutionary strata’. This suggested that at least four events had punctuated human sex chromosome evolution, with each event suppressing X–Y crossing over in one stratum without grossly disturbing gene order in the X chromosome.

We re-analysed this published information and combined the results with K_s and map location data for 12 additional X–Y gene pairs, thus compiling data on 31 X–Y pairs in all (Supplementary Table 5). In each of 27 pairs, the Y member is an X-degenerate gene

or pseudogene. The other four pairs include two in which the Y member is an X-transposed gene and two in which the Y members are ampliconic gene families.

Among all X-degenerate pairs, and the two ampliconic pairs, the previously reported correlation between age (K_s) and X map position is readily apparent, with age increasing from the distal short arm to the long arm of the X chromosome (Fig. 7). Furthermore, as observed in the earlier study, the order of the homologous genes in the MSY appears to be scrambled with respect to K_s (Supplementary Fig. 9). These observations, together with the

Table 2 MSY genes and gene families demonstrated or hypothesized to encode proteins

MSY sequence class	Gene symbol	Gene name	Number of copies†	Tissue expression	X-linked homologue	Autosomal homologue
X-transposed	<i>TGIF2LY*</i>	TGF (beta)-induced transcription factor 2-like Y	1	Testis	<i>TGIF2LX</i>	–
	<i>PCDH11Y</i>	Protocadherin 11 Y	1	Fetal brain, brain	<i>PCDH11X</i>	–
Total			2			
X-degenerate	<i>SRY</i>	Sex determining region Y	1	Predominantly testis	<i>SOX3</i>	–
	<i>RPS4Y1</i>	Ribosomal protein S4 Y isoform 1	1	Ubiquitous	<i>RPS4X</i>	–
	<i>ZFY</i>	Zinc finger Y	1	Ubiquitous	<i>ZFX</i>	–
	<i>AMELY</i>	Amelogenin Y	1	Teeth	<i>AMELX</i>	–
	<i>TBL1Y*</i>	Transducin (beta)-like 1 protein Y	1	Fetal brain, prostate	<i>TBL1X</i>	–
	<i>PRKY</i>	Protein kinase Y	1	Ubiquitous	<i>PRKX</i>	–
	<i>USP9Y</i>	Ubiquitin-specific protease 9 Y	1	Ubiquitous	<i>USP9X</i>	–
	<i>DBY</i>	Dead box Y	1	Ubiquitous	<i>DBX</i>	–
	<i>UTY</i>	Ubiquitous TPR motif Y	1	Ubiquitous	<i>UTX</i>	–
	<i>TMSB4Y</i>	Thymosin (beta)-4 Y	1	Ubiquitous	<i>TMSB4X</i>	–
	<i>NLGN4Y</i>	Neurologin 4 isoform Y	1	Fetal brain, brain, prostate, testis	<i>NLGN4X</i>	–
	<i>CYorf15A*</i>	Chromosome Y open reading frame 15A	1	Ubiquitous	<i>CXorf15</i>	–
	<i>CYorf15B*</i>	Chromosome Y open reading frame 15B	1	Ubiquitous	<i>CXorf15</i>	–
	<i>SMCY</i>	SMC (mouse) homologue, Y	1	Ubiquitous	<i>SMCX</i>	–
	<i>EIF1AY</i>	Translation initiation factor 1A Y	1	Ubiquitous	<i>EIF1AX</i>	–
	<i>RPS4Y2*</i>	Ribosomal protein S4 Y isoform 2	1	Ubiquitous	<i>RPS4X</i>	–
Total			16			
Ampliconic	<i>TSPY</i>	Testis-specific protein Y	~35	Testis	–	–
	<i>VCY</i>	Variable charge Y	2	Testis	<i>VCX</i>	–
	<i>XKRY</i>	XK related Y	2	Testis	–	–
	<i>CDY</i>	Chromodomain Y	4	Testis	–	<i>CDYL</i>
	<i>HSFY*</i>	Heat shock transcription factor Y	2	Testis	–	–
	<i>RBMY</i>	RNA-binding motif Y	6	Testis	<i>RBMX</i>	–
	<i>PRY</i>	PTP-BL related Y	2	Testis	–	–
	<i>BPY2</i>	Basic protein Y 2	3	Testis	–	–
	<i>DAZ</i>	Deleted in azoospermia	4	Testis	–	<i>DAZL</i>
Total			~60			
Grand total			~78			

See Supplementary Table 1 for a more detailed version of this Table, incorporating references and GenBank accession numbers.
* Genes first reported in this study.
† Excluding pseudogenes.

Sequence class	Defining characteristics	Evolutionary origins	Distribution	Aggregate length (Mb)	No. of coding genes	No. of non-coding transcription units	No. of transcription units per Mb	Nucleotides in interspersed repeats (%)
X-transposed	99% identity to X	Single transposition from X	2 blocks on Yp	3.4	2	0	0.6	60
X-degenerate	Single-copy gene or pseudogene homologues of X-linked genes	Relics of ancient autosomes from which X and Y evolved	8 blocks on Yp and Yq	8.6	16, most expressed widely	4	2.2	57
Ampliconic	Lengthy similarity to other MSY sequences	Acquired from diverse sources, then amplified	7 blocks on Yp and Yq	10.2	60 (in 9 families), expressed mainly or only in testes	74 (9 single-copy; 65 in 15 families), expressed mainly or only in testes	13.3	36

Figure 4 Three sequence classes in the MSY euchromatin. Colour scheme as in Fig. 2.

earlier arguments of Lahn and Page, suggest three conclusions. First, all MSY genes and pseudogenes identified here as X-degenerate seem to be products of a single molecular evolutionary process: the region-by-region suppression of crossing over in ancestral autosomes, with subsequent differentiation of the Y from the X chromosome (Fig. 6). Second, at least two of the MSY's ampliconic gene families, *VCY* and *RBMV*, also originated in this manner, but subsequently acquired the characteristics of ampliconic sequences (Fig. 6; for independent evidence concerning *RBMV* see refs 27 and 28). Third, as previously hypothesized, inversions in the Y chromosome may have suppressed crossing over with the X chromosome.

X-transposed genes as exceptions

A very different evolutionary model accounts for the X-transposed genes, as confirmed by our K_s analysis. If, as hypothesized, these MSY genes are the result of a single, recent transposition from the X chromosome (Fig. 6), then the K_s values of the two X-transposed X–Y gene pairs should be similar to each other but much lower than the K_s values of the nearby (X-degenerate) pairs in the X-chromosome long arm. This prediction is met (Fig. 7). The two X-transposed X–Y gene pairs seem to be orders of magnitude younger than the ancient pairs (group 1 in Fig. 7) among which they are physically situated in the X chromosome.

Table 3 MSY palindromes				
Palindrome	Arm length (kb)	Arm-to-arm identity (%)	Spacer length (kb)	Palindrome span (kb)
P1	1,450	99.97	2.1	2,902
P1.1*	9.9	99.95	3.9	24
P1.2*	9.9	99.95	3.9	24
P2	122	99.97	2.1	246
P3	283	99.94	170	736
P4	190	99.98	40	419
P5	496	99.98	3.5	996
P6	110	99.97	46	266
P7	8.7	99.97	12.6	30
P8	36	99.997	3.4	75
Total span (kb)				5,670

*Palindromes P1.1 and P1.2 are located within, respectively, the distal and the proximal arms of palindrome P1 (Fig. 2).

Blurred boundaries

Our observations differ from those of Lahn and Page in that the boundaries between X–Y gene groups 2 and 3, and between groups 3 and 4, now seem less distinct (Fig. 7; compare with Fig. 2 in ref. 26). Whereas our present observations could be interpreted as evidence that suppression of X–Y crossing over evolved in more than four steps, such a conclusion would be premature. The apparent overlaps

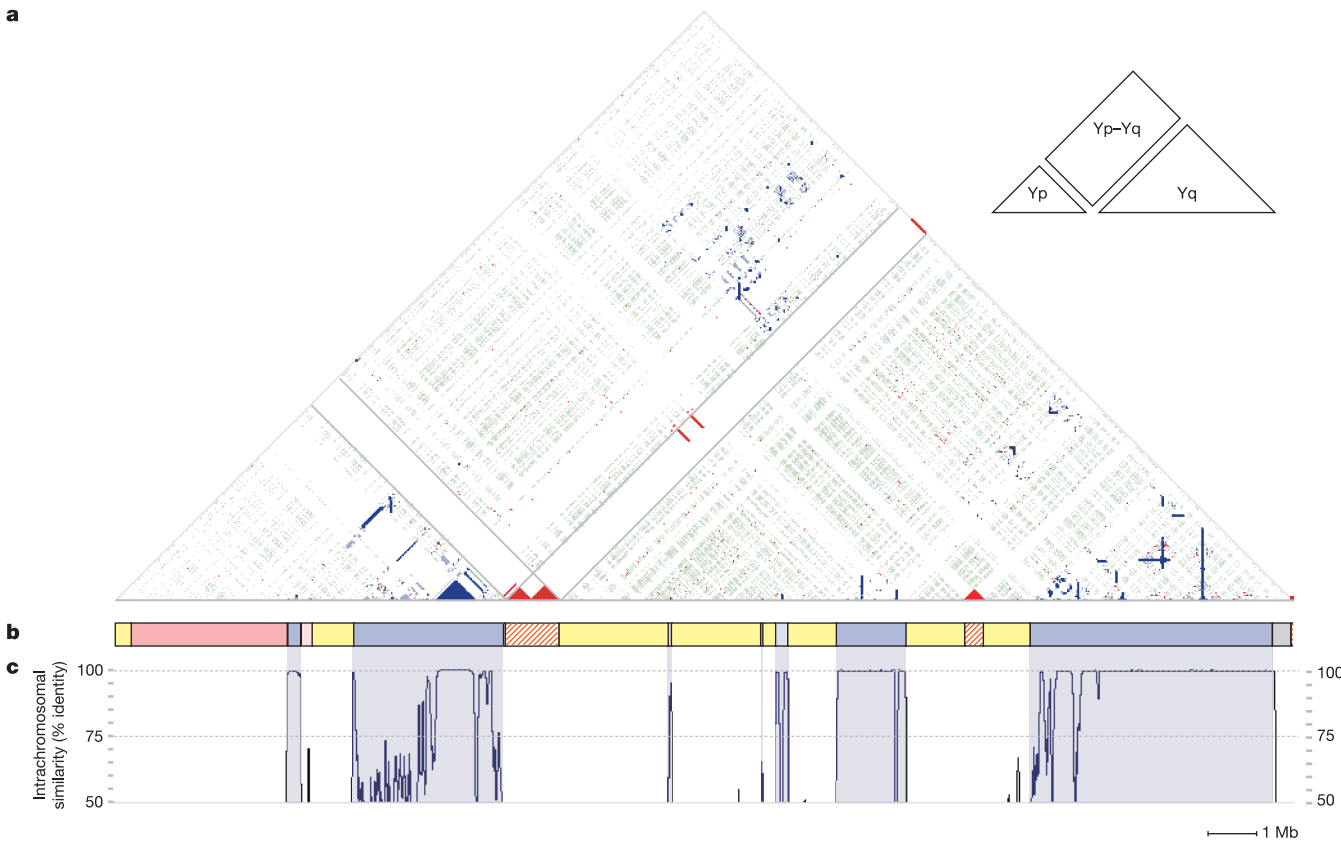


Figure 5 Sequence similarities within the MSY. **a**, Triangular dot plot in which the MSY's sequence is compared to itself. Within the plot, each dot represents a match of >65% within a window of 2,000 nucleotides. Green dots represent matches of this quality between LINE1 elements; red dots represent matches between heterochromatic sequences; blue dots represent matches between all other sequences. Direct repeats appear as horizontal lines, inverted repeats as vertical lines, and palindromes as vertical lines that nearly intersect the baseline. Long arrays of tandem repeats appear as pyramids. The inset indicates that the large triangular plot contains two smaller triangles (one revealing sequence similarities within Yp, and one revealing similarities within Yq)

and a rectangle (revealing similarities between Yp and Yq). **b**, MSY schematic, as in Fig. 1b. **c**, Plot of intrachromosomal sequence similarity, which serves to identify ampliconic sequences (blue). Using a 50-kb sliding window and 1-kb steps, each MSY euchromatic sequence was compared to all other available MSY euchromatic sequences. (Long interspersed repeats were excluded before analysis.) At each point along the length of the MSY, the highest sequence similarity (expressed as per cent nucleotide identity) was identified. All such values >50% are shown. An expanded version of this plot is shown in Fig. 2f.

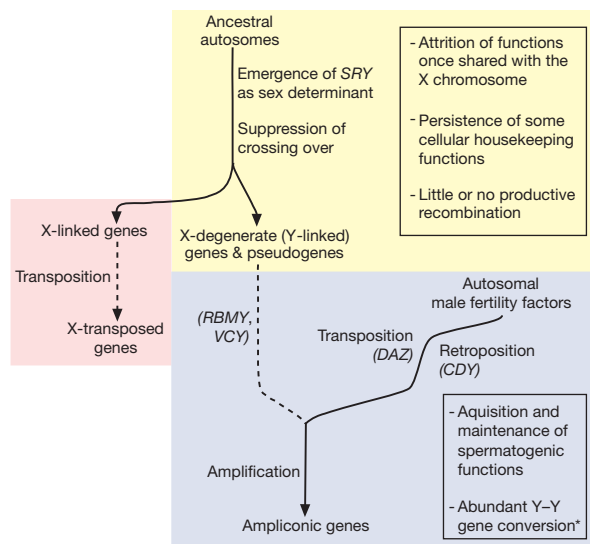


Figure 6 Molecular evolutionary pathways and processes that gave rise to genes in three MSY euchromatic sequence classes. X-degenerate genes and pseudogenes (yellow background) derived from an autosomal pair that was ancestral to both the X and Y chromosomes (and that was enlarged by subsequent fusion with other autosomes or autosomal segments⁵⁰). X-transposed genes (pink background) derived from X-linked genes, which in turn derived from the ancestral autosomal pair. Ampliconic genes (blue background) were derived through three converging processes: amplification of X-degenerate genes (for example, *RBMY*, *VCY*); transposition and amplification of autosomal genes (*DAZ*); and retroposition and amplification of autosomal genes (*CDY*). Boxes enumerate dominant themes in X-degenerate (yellow) and ampliconic (blue) gene evolution. The asterisk indicates that Y–Y gene conversion is apparently common in the 61% of ampliconic sequences that exhibit intrachromosomal identities of $\geq 99.9\%$.

between groups could be artefacts of local errors in ordering X-linked genes, these regions not yet having been fully sequenced, or simply of large standard errors for some K_s estimates (Fig. 7). Some changes in local gene order in the X chromosome may also have occurred during its evolution. Another potentially confounding factor is X–Y gene conversion, which would depress K_s values and estimated ages for gene-converted X–Y pairs. Gene conversion depends on high sequence similarity, and thus one might expect any such effect to be greater among the younger X–Y pairs, in groups 3 and 4. Indeed, comparisons of X and Y genomic sequences suggest that the *VCX/Y* pair and 3' portions of the *KAL1/P* pair (both pairs in group 4) have engaged in extensive gene conversion (Supplementary Fig. 10), depressing their K_s values below those of the 5' portion of the *KAL1/P* pair and of other group 4 pairs (Fig. 7).

A map of male-specific ages

Having examined the evolutionary ages of all 31 X–Y gene pairs, we used them to anchor an evolutionary map of the modern human MSY. The map displays the male-specific ages of many sequence segments (Fig. 8). Here, male-specific age is the estimated number of years that have passed since sequences ancestral to that segment were incorporated into the MSY (having previously been autosomal, pseudoautosomal, or X-linked). We estimated the age of each gene or segment using Lahn and Page's methods that combined K_s analysis (Supplementary Table 5) with comparative gene mapping data from other mammals. The resulting estimated ages are graphed on a logarithmic scale to accommodate a range that extends from approximately 4 million years (the X-transposed sequences; the youngest known sequences in the MSY) to approximately 300 million years (*SRY*, the sex determinant and arguably the oldest gene in the MSY).

As can be seen in Fig. 8, the MSY euchromatin is an elaborate patchwork of sequences of diverse male-specific ages. The result of a single, recent transposition from the X-chromosome, the MSY's X-transposed sequences are homogeneously youthful. The

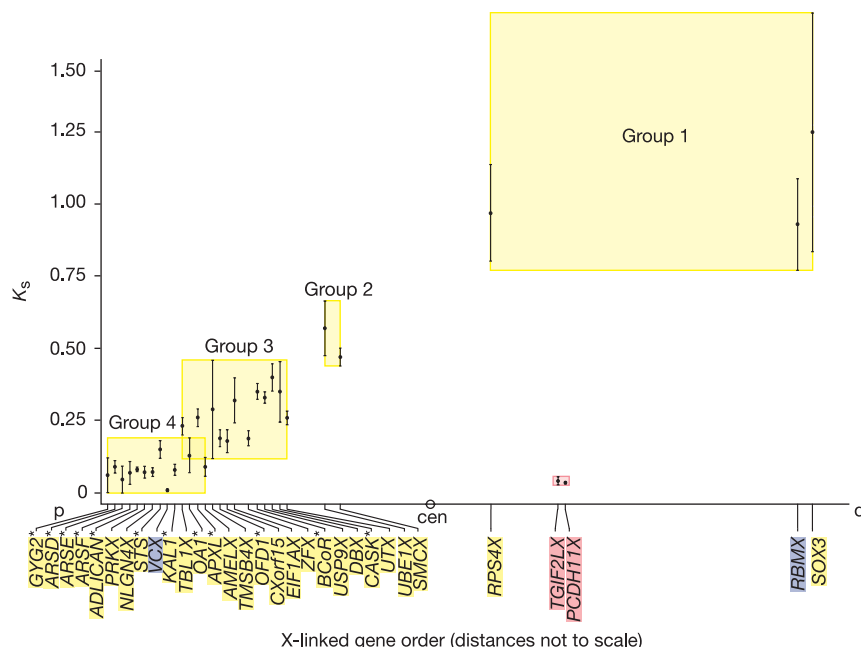


Figure 7 Plot of K_s (Supplementary Table 5) versus X-linked gene order for 31 X–Y gene (or gene/pseudogene) pairs. Colour highlighting of X-linked gene names indicates whether Y homologues are X-degenerate (yellow), ampliconic (blue) or X-transposed (pink). Within the plot, four yellow rectangles denote four previously defined 'evolutionary

strata', or groups of genes²⁶; a small pink rectangle highlights two X-transposed genes. Genes in the X chromosome are ordered according to the NCBI sequence assembly of November 2002; distances between genes are not drawn to scale. Standard errors for K_s values are shown.

sequences of both the X-degenerate and ampliconic classes are much older, and they display a wide range of male-specific ages (Fig. 8). As we will argue, it is in comparing and contrasting these two chronologically diverse classes that the central themes of MSY evolution and function are revealed most clearly.

Evolutionary dynamics of X-degenerate and ampliconic sequences

To appreciate the evolutionary dynamics of these two sequence classes, we need to consider both their similarities and differences. In many senses, the X-degenerate and ampliconic sequences together dominate the euchromatic MSY. The X-degenerate and ampliconic classes are physically intermingled in the MSY, and they are comparably large, constituting, respectively, 38% and 45% of the MSY's euchromatic sequences (Fig. 1 and Supplementary Table 6). Together, these two sequence classes carry all but two of the MSY's 78 known protein-coding transcription units (Table 2). The X-degenerate and ampliconic classes display comparable diversities of male-specific ages, from tens to hundreds of millions of years (Fig. 8). This implies that X-degenerate and ampliconic sequences evolved in parallel, as parts of a single DNA molecule, for as much as 300 million years. Moreover, we infer that the X-degenerate and ampliconic sequences evolved under similar, unusual circumstances: both were transmitted exclusively through the male germ line, and neither participated in meiotic crossing over with a homologous counterpart. However, a number of marked structural and functional differences between these two sequence classes

suggest that they followed different evolutionary trajectories. Palindromes are prevalent in ampliconic sequences. The density of transcription units is much higher and the density of interspersed repeats is much lower in ampliconic than in X-degenerate sequences (Fig. 1). The two sequence classes also diverge starkly with respect to gene-expression patterns. Most X-degenerate genes are expressed widely throughout the body, and many are probably involved in cellular housekeeping activities that are critical in both males and females. In contrast, most ampliconic genes are expressed predominantly or exclusively in testes, where they probably function in spermatogenesis.

Decay in the absence of sexual recombination

The X-degenerate sequences are adequately explained by the prevailing theory of sex chromosome evolution, which states that as the X and Y chromosomes evolved from an autosomal pair, the X chromosome maintained most of its ancestor's genes whereas the Y chromosome lost them^{5,24–26}. Our findings support the two major premises of this theory: the evolutionary genetic benefits of sexual recombination through meiotic crossing over, and the deleterious consequences of its absence. According to this theory, most ancestral genes remained functionally intact in the X chromosome, where the benefits of crossing over (in females) continued. In the Y chromosome, in contrast, the shutting down of X–Y crossing over during evolution triggered a monotonic decline in gene function. This model is corroborated by the presence, in the MSY's X-degenerate sequences, of decayed, intron-bearing pseudogenes of

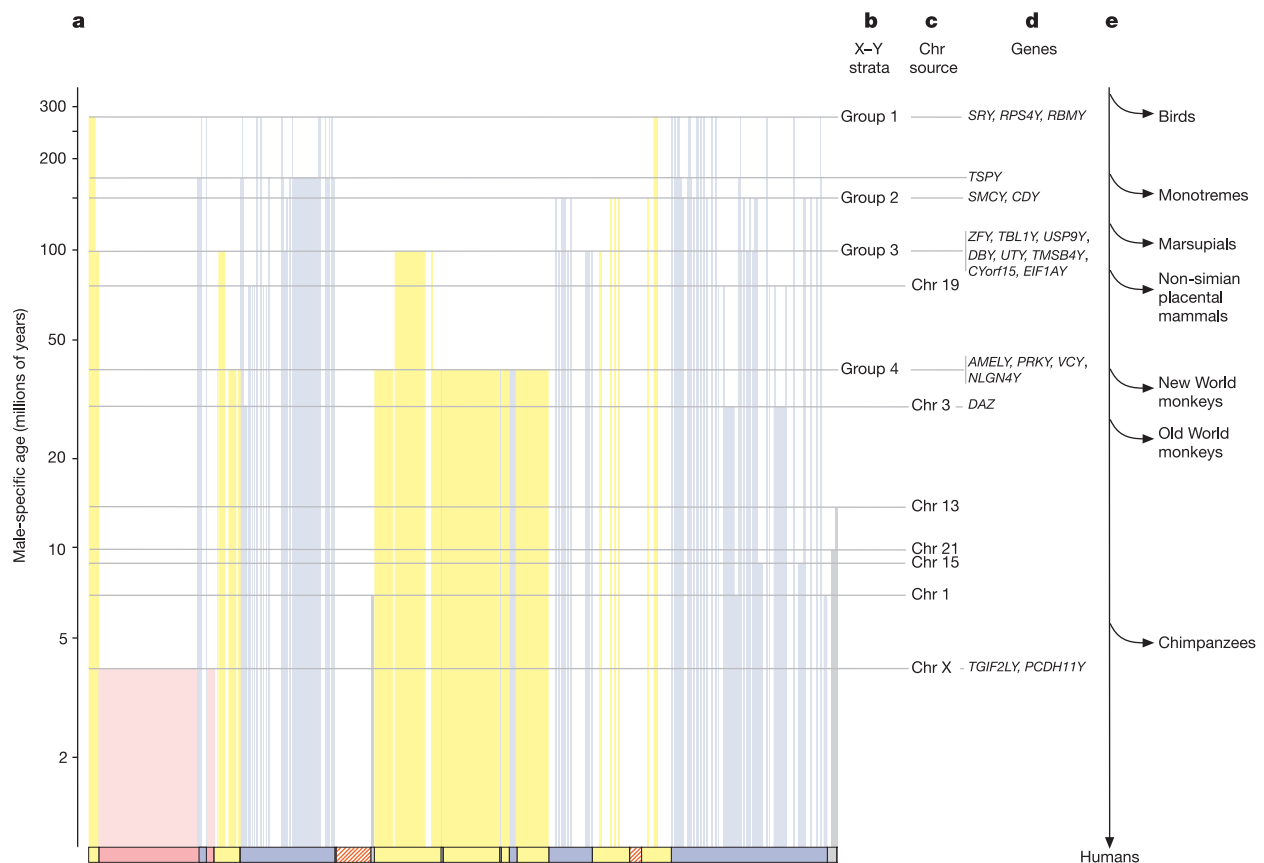


Figure 8 Evolutionary map of the MSY. At the bottom is an MSY schematic, as in Fig. 1b. Coloured rectangles extending above this schematic depict the estimated male-specific ages of the corresponding segments of the modern MSY. These ages are plotted on a logarithmic scale (a). b, X–Y strata 1, 2, 3 and 4 (ref. 26 and Fig. 7). c, The chromosomes (more properly, the modern human orthologues of the chromosomes) from which the

indicated X-transposed or ampliconic sequences apparently arose through transposition, during evolution. d, MSY genes that apparently arose at the indicated times. e, Approximate times of divergence between the human and certain other vertebrate lineages. The methods used to estimate the male-specific ages of each of the sequences and genes shown are listed in Supplementary Table 9.

13 different X-linked genes (Supplementary Table 3). Presumably, many hundreds of other X-homologous genes were deleted outright from the evolving MSY, leaving no trace in the DNA sequence of the modern human MSY. Seen in this light, the 16 protein-coding genes in the modern MSY's X-degenerate sequences (Table 2 and Fig. 3) appear as rare examples of persistence in the absence of sexual recombination.

Acquisition and conservation of spermatogenic functions

This evolutionary model of the Y chromosome as a decaying X chromosome, however, provides no explanation for central characteristics of the MSY's ampliconic sequences, including testis-specific gene expression, near-perfect palindromes, and an abundance of autosomal (as well as X-chromosomal) sequence similarities. To account for these characteristics, we propose that the MSY acquired, and evolved a means of conserving, genes that specifically enhanced male fertility.

Unlike the X-degenerate sequences, all of which trace to the MSY's shared ancestry with the X chromosome, the ampliconic sequences evolved from a great variety of genomic sources, and by a diversity of molecular mechanisms (Fig. 6). As mentioned previously, the ampliconic genes *VCY* and *RBMV* were, similar to the X-degenerate genes, derived from common ancestors of the X and Y chromosomes^{27,28}. In contrast, the *DAZ* genes arose, during primate evolution, by transposition and subsequent amplification of an autosomal transcription unit, *DAZL*, which still exists on human chromosome 3 (ref. 29). Indeed, systematic analysis of MSY/autosomal similarities suggests that a series of autosomal transpositions contributed to the MSY's ampliconic sequences during primate evolution (Fig. 8; see also ref. 13). Yet another molecular mechanism accounts for the *CDY* genes, which arose by retroposition (and subsequent amplification) of a processed messenger RNA derived from an autosomal gene³⁰. This retroposition event was previously thought to have occurred during primate evolution, but our present *K_s* analysis indicates a much older date, probably before the lineages of marsupials and placental mammals diverged (Fig. 8; see also Supplementary Table 5).

Despite the wide variety of genomic sources and molecular evolutionary mechanisms that gave rise to the ampliconic genes, they all came to exist in the MSY in multiple, nearly identical copies, and they evolved remarkably uniform patterns of tissue expression. Indeed, detailed studies of several ampliconic gene families have revealed that they are expressed predominantly or exclusively in one cell lineage: the spermatogenic cells of the testis. What accounts for this convergence of evolutionary outcomes? The genesis of XY sex chromosomes during mammalian evolution, and specifically the emergence of a male-specific domain, created a genomic niche where selection could operate to enhance male germ-cell development. Amplification of the testis genes might have enhanced sperm production through high levels of expression. However, in a region

devoid of crossing over, amplification might also have allowed another type of homologous recombination, gene conversion, to emerge as a means of conserving gene function.

Abundant Y–Y gene conversion in ampliconic regions

Gene conversion is the non-reciprocal transfer of sequence information from one DNA duplex to another³¹. This type of genetic recombination has been studied most extensively in fungi, where it was originally demonstrated to occur between chromosome homologues, or at lower frequency between sister chromatids, in meiosis. It was later shown that gene conversion could also occur between duplicated sequences on a single chromosome, and in mitosis³². Here we will argue that gene conversion (non-reciprocal recombination) is as frequent in the MSY as crossing over (reciprocal recombination) is in ordinary chromosomes.

Specifically, two major findings provide evidence that gene conversion occurs routinely in 30% of the MSY euchromatin, including nearly all of the MSY's testis-specific gene families. The accompanying study⁷ reports the identification and sequencing of chimpanzee Y-linked orthologues of human MSY palindromes and establishes that gene conversion between palindrome arms has occurred in both the human and chimpanzee lineages, and has continued to occur in human populations. Here we report that these palindromes are representative of a large, discrete fraction of MSY sequences, all of which bear at least 99.9% identity to other MSY sequences. These findings suggest that the entire fraction is subject to frequent gene conversion.

Above we described calculations of percentage nucleotide identity between all pairs of known MSY sequences. We defined and mapped the ampliconic regions by reporting, at each point along the length of the MSY euchromatin, the highest percentage identity to other MSY sequences (intrachromosomal similarity; Fig. 5). To view this data from another perspective, we electronically fractionated all MSY sequences according to intrachromosomal similarity. As seen in Fig. 9a, 30% of MSY euchromatic sequences display intrachromosomal identities of 99.9–100%. As intrachromosomal identity declines below 99.9%, the fractional representation of MSY sequences drops abruptly. Thus, the sequences displaying intrachromosomal identities of ≥99.9% represent a large and distinct subset of the MSY euchromatin.

This ≥99.9% subset comprises the eight palindromes as well as large portions of the IR2 and IR3 inverted repeats described above (Figs 2 and 3). Indeed, nearly all of the ≥99.9% sequences exist as pairs in inverted orientation. Thus, the MSY palindromes in which gene conversion has been demonstrated⁷ are typical and representative of the ≥99.9% fraction. We extrapolate that nearly all of the ≥99.9% fraction is engaged in gene conversion on a routine basis, resulting in a degree of identity among MSY's inverted sequence pairs that rivals that of two autosomal homologues, or alleles, chosen at random from the human population^{15,33}.

Table 4 MSY testis gene family members in regions exhibiting ≥99.9% or <99.9% intrachromosomal identity

Gene family	Number of genes		Number of pseudogenes	
	Regions ≥99.9% intrachromosomal identity	Regions <99.9% intrachromosomal identity	Regions ≥99.9% intrachromosomal identity	Regions <99.9% intrachromosomal identity
<i>VCY</i>	2	0	0	0
<i>XKRY</i>	2	0	6	0
<i>CDY</i>	4	0	17	4
<i>HSFY</i>	2	0	0	0
<i>RBMV</i>	6	0	6	17
<i>PRY</i>	2	0	4	0
<i>BPY2</i>	3	0	Many	Many
<i>DAZ</i>	4	0	0	0
Total	25	0	>33	>21

The *TSPY* genes, the only MSY genes found in long tandem arrays, are excluded from this analysis.

Two modes of productive recombination in the human Y chromosome

Combined with previous discoveries in the pseudoautosomal regions, the present findings imply that two modes of homologous recombination occur regularly in the human Y chromosome. First, there is crossing over with the X chromosome in the pseudoautosomal regions (aggregate length 3.0 Mb) (Supplementary Note 6). Second, there is Y–Y gene conversion in the $\geq 99.9\%$ regions (aggregate length 6.1 Mb) dispersed throughout the MSY (Fig. 9b)⁷. We refer to both routine modes of Y chromosome recombination as ‘productive’ to distinguish them from the relatively rare, aberrant recombination events (typically Y–Y or X–Y) that perturb sex differentiation or fertility and thereby diminish the reproductive fitness of affected individuals.

Genetic mapping studies have shown that, typically, one X–Y crossover occurs per generation in the pseudoautosomal regions (Supplementary Note 7). As described in the accompanying report⁷, steady-state calculations suggest that, on average, multiple Y–Y gene conversion events take place per generation in the MSY. Thus, most homologous recombination events in the Y chromosome probably occur in the MSY.

In recent years, we and other investigators have referred to the MSY as the NRY, or ‘non-recombining region of the Y chromosome’. This usage reflected both awareness that productive X–Y crossing over did not occur in the MSY, and ignorance of the Y–Y gene conversion that is apparently commonplace there. We now refer to the NRY as the MSY, or ‘male-specific region of the Y chromosome’, because it is recombinogenic and unique to males.

Gene conversion and the MSY's testis gene families

Examination of the MSY's testis gene families provides additional insight into the potential biological significance of the $\geq 99.9\%$ fraction and the gene conversion associated with it. Eight of the

MSY's nine identified testis gene families have members in the palindromes or inverted repeats that comprise the $\geq 99.9\%$ fraction just described. (The exceptional family is *TSPY*, most of whose members are found in a long tandem array.) Many of these family members are intact gene copies, but others are apparent pseudogenes with disrupted splice sites or reading frames. For each of the eight testis gene families, we counted the numbers of intact and pseudogene copies, both within and without the $\geq 99.9\%$ fraction (Table 4). Whereas large numbers of pseudogenes are present both inside and outside the $\geq 99.9\%$ fraction, the intact gene copies, 25 in all, are located exclusively in the $\geq 99.9\%$ fraction.

Thus, there is an evident association of intact testis genes with near-identical inverted sequence pairs that undergo gene conversion. What is the biological significance of this association? We envision two possibilities, which are not mutually exclusive. First, we note that in all cases examined so far, expression of these testis-specific gene families has been found to be limited to or most pronounced in cells of the spermatogenic lineage—in germ cells. Perhaps these near-identical sequence pairs are transcriptionally active in germ cells because there they generate cruciforms or other unusual chromatin configurations. Second, the occurrence of MSY gene pairs that are subject to frequent gene conversion might provide a mechanism for conserving gene functions across evolutionary time in the absence of crossing over.

Implications for future studies

We anticipate that the nucleotide sequence reported here, and the methods with which it was obtained, will find many applications in human biology and beyond.

Comparisons with other human Y chromosomes

The sequence of one man's MSY, as reported here, provides a point

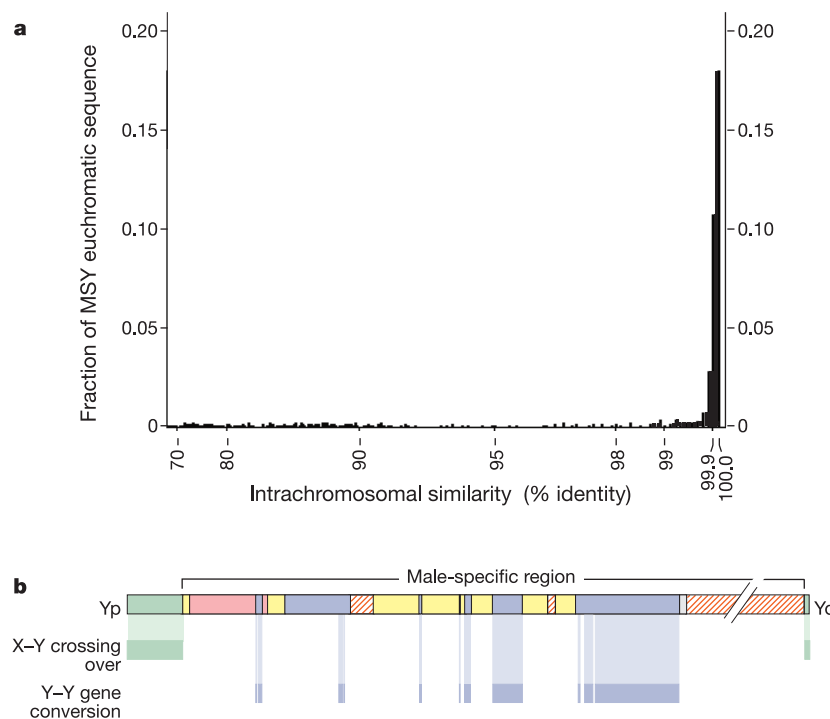


Figure 9 MSY sequences exhibiting $\geq 99.9\%$ intrachromosomal identity probably undergo Y–Y gene conversion. **a**, Electronic fractionation of MSY euchromatic sequences according to intrachromosomal similarity (per cent identity to other MSY sequences), plotted on a logarithmic scale. Values $< 70\%$ are not shown. **b**, Sites of productive recombination in the Y chromosome. Shown at the top is a schematic representation of

the entire Y chromosome, including the pseudoautosomal regions (green). The pseudoautosomal regions are sites of frequent X–Y crossing over. Within the MSY's ampliconic sequences are many sites of apparently frequent Y–Y gene conversion; all of these sites display intrachromosomal identities of $\geq 99.9\%$.

of departure for systematic, comprehensive characterization of MSY sequence variation in human populations. The MSY's unique characteristics—male specificity, no crossing over and abundant gene conversion—suggest that its sequence variation might differ markedly from that of ordinary human chromosomes. Already the availability of MSY sequence information in public databases has accelerated the emergence of MSY sequence variation as a powerful tool in reconstructing the patrilineal origins of modern human populations^{11,34}.

Comparisons (or lack of) with other species

Little is known about the DNA sequences of Y chromosomes in other animals or plants, and thus it is not possible at present to compare systematically the human MSY with that of any other species. Both the *Drosophila* and mouse Y chromosomes contain genes required for spermatogenesis, but meagre Y chromosome sequence data is available in either species. In *Drosophila*, the sequences of autosomes and the X chromosome were assembled from whole-genome shotgun data. Unfortunately, this shotgun analysis was insufficient to assemble much Y chromosome sequence^{35,36}, confirming prior suspicions that, in *Drosophila* as in humans, the Y chromosome poses special challenges. In the mouse, a draft sequence of the female genome is available³⁷, but systematic efforts to sequence the male-specific region of the Y chromosome have yet to be initiated. If undertaken, Y chromosome sequencing projects in *Drosophila*, mouse and other species are likely to encounter special technical hurdles, but they are also likely to yield entirely unforeseen biological insights, as was the case here for the human MSY. The availability of human MSY sequence has already enabled new tests and rekindled debate of Haldane's hypothesis that mutations in the male germ line greatly outnumber those in the female germ line (Supplementary Note 8). This debate will surely be fuelled by sequencing of other primate and mammalian Y chromosomes.

Methods for sequencing difficult genomic regions

Our strategy of iterative mapping and sequencing was laborious but essential. Two faster, less costly strategies have been used recently in sequencing large genomes: whole-genome shotgun analysis^{15,35} and sequencing a tiling path of mapped clones (ref. 14 and Supplementary Note 9). Neither of these sequencing strategies would have yielded a coherent picture of the MSY. This is especially true of the MSY's ampliconic regions, and most particularly the 30% of the MSY euchromatin (including the eight palindromes) exhibiting intrachromosomal similarities of $\geq 99.9\%$. Large amplicons like those described here are not unique to the MSY, but as in the MSY, they have proven to be formidable obstacles to whole-genome methods^{38,39}. The iterative mapping and sequencing strategy used here should be considered by genome scientists wishing to determine the structure and sequence of amplicon-rich regions of human autosomes, the X chromosome and other genomes.

The medical relevance of the MSY

Propelled by advances in MSY genomics, the biomedical significance of the MSY has begun to surface in recent years, with evidence of roles in such diverse processes as gonadal sex determination, skeletal growth, germ-cell tumorigenesis and graft rejection⁶. Two research areas that should benefit from the present MSY sequence and gene catalogue are of particular note. First, one of the most common chromosomal disorders of girls and women is Turner syndrome, classically associated with a 45,X (X0) karyotype. Haploinsufficiency of particular genes common to the X and Y chromosomes may be responsible for somatic features of the syndrome^{40–42}. In most cases, the molecular identity of these Turner genes remains to be determined. One or more Turner genes are likely to be found within the catalogue of X-degenerate genes (and their X-linked homologues; see Table 2).

A highly active area of MSY research explores spermatogenesis and the genetic basis of male infertility. MSY deletions have emerged as the most common of the known genetic causes of spermatogenic failure in human populations^{13,43–46}. The availability of MSY sequence has already begun to transform our understanding, enabling investigators to precisely define four distinct classes of recurrent MSY deletions causing spermatogenic failure, identify the MSY genes absent as a result of these deletions (typically members of testis-specific families), and demonstrate that most such deletions are the result of homologous recombination between near-identical amplicons^{13,43–46}. Thus, the ampliconic structures that may help preserve testis gene function across evolutionary time (through gene conversion) also put individuals at risk of spermatogenic failure (again, through homologous recombination).

Genetic and biological differences between males and females

It is commonly stated that the genomes of two randomly selected members of our species exhibit 99.9% nucleotide identity. In reality, this statement holds only if one is comparing two males, or two females. If one compares a female with a male, the second X chromosome (160 Mb, or roughly 3% of the diploid DNA content) is replaced by the largely dissimilar Y chromosome (60 Mb, or 1% of the diploid DNA content). This common substitution of the Y chromosome for the second X chromosome dwarfs all other DNA polymorphism in the human genome. In decades past, and with the important exception of X-linked recessive diseases, biologists often judged this genomic dimorphism to be of limited functional consequence, especially because of inactivation of the second X chromosome in females and the presumed paucity of genes in the Y chromosome. Now we must begin to reconsider this position, given the unanticipated number and variety of MSY genes, many of which are expressed throughout the body, and the fact that many X-linked genes are expressed from both X chromosomes in female cells⁴⁷. The present sequence of the MSY, and the emerging sequence of the X chromosome, offer the near prospect of a comprehensive catalogue of genetic and sequence differences between human males and females. Translating this knowledge into an understanding of the myriad differences between the sexes in anatomy, physiology, cognition, behaviour and disease susceptibility presents a monumental challenge, but surely one of broad significance and interest. □

Methods

Iterative mapping and sequencing

The method of iterative mapping and sequencing used here has been described^{10,13}. All MSY BACs selected for sequencing were isolated from the RPCI-11 library⁴⁸, with the exception of 11 clones (nine spanning the *AZF*a region¹², and two used to narrow gaps¹⁰) from the CITB and CITC libraries. We made frequent use of publicly available BAC-end sequences as a source of markers during the final stages of map construction⁴⁹. Two gaps were closed by long-range PCR; see Supplementary Fig. 11.

Unfortunately, no cell line is available from the donor of the RPCI-11 BAC library. Thus, to confirm the large-scale organization of MSY sequences reported here, we PCR-amplified the inner and outer boundaries of all palindromes in ten men with genetically diverse Y chromosomes (PCR primers in Supplementary Table 8). We sequenced all resulting products. These experiments confirmed that each palindrome boundary is present in the great majority of human Y chromosomes.

Intrachromosomal sequence similarity

Analyses of intrachromosomal similarity were performed using custom Perl code. This code used BLAST (<http://blast.wustl.edu>) to compare all 5-kb sequence segments, in 2-kb steps, to the entire remainder of the MSY sequence.

Interspersed repeats

We electronically identified interspersed repeats with RepeatMasker (<http://repeatmasker.genome.washington.edu>).

Homology to other chromosomes

To identify sequence similarities to other human chromosomes, we conducted BLAST searches against GenBank databases with the sequence of each MSY clone. Interspersed repeats and low-complexity regions were masked using RepeatMasker. To experimentally verify the chromosomal origins of sequences similar to the MSY, we designed STSs from

those sequences and assayed them against the NIGMS human/rodent somatic cell hybrid mapping panels 1 and 2 (NIGMS Human Genetic Cell Repository, <http://locus.umd.edu/nigms/maps/mapping.html>).

Identification of new genes and transcription units

We identified potential transcripts from three sources: (1) BLAST matches to cDNA sequences (EST or full length). We pursued matches where the cDNA sequence showed evidence of polyadenylation or splicing, or where there were multiple matching cDNA sequences. (2) BLAST matches to fragments of putative MSY transcripts that had been cloned by cDNA selection of testis cDNA against a flow-sorted, genomic Y-chromosome library¹⁹. (3) GenomeScan¹⁸ predictions in the NCBI annotation of Y-chromosome contigs. We then tested for transcription by RT-PCR as previously described¹³.

Chromosomal FISH

One- or two-colour FISH to human chromosomes was performed as previously described⁹.

Calculation of K_s and K_a

We calculated the numbers of synonymous substitutions per synonymous site (K_s) and of non-synonymous substitutions per non-synonymous site (K_a) as follows. We used FASTA (<ftp://ftp.virginia.edu/pub/fasta>) to align the pairs of coding sequences in Supplementary Table 5. For non-transcribed MSY pseudogenes, we used FASTA to align the genomic sequence of pseudogene exons to the corresponding transcribed coding sequence (Supplementary Table 5 and File 3). Then, as is standard practice, insertions/deletions were manually removed from the alignments. We calculated K_s and K_a for these alignments using the diverge function in the Wisconsin Package (Version 10.2, Genetics Computer Group).

Received 7 March; accepted 8 April 2003; doi:10.1038/nature01722.

1. Painter, T. S. The Y-chromosome in mammals. *Science* **53**, 503–504 (1921).
2. Stern, C. The problem of complete Y-linkage in men. *Am. J. Hum. Genet.* **9**, 147–166 (1957).
3. Jacobs, P. A. & Strong, J. A. A case of human intersexuality having a possible XXY sex determining mechanism. *Nature* **183**, 302–303 (1959).
4. Ford, C. E., Miller, O. J., Polani, P. E., de Almeida, J. C. & Briggs, J. H. A sex-chromosome anomaly in a case of gonadal dysgenesis (Turner's syndrome). *Lancet* **1**, 711–713 (1959).
5. Ohno, S. *Sex Chromosomes and Sex-linked Genes* (Springer, Berlin, 1967).
6. Vogt, P. H. *et al.* Report of the third international workshop on Y chromosome mapping 1997. *Cytogenet. Cell Genet.* **79**, 1–20 (1997).
7. Rozen, S. *et al.* Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature* **423**, 873–876 (2003).
8. Foote, S., Vollrath, D., Hilton, A. & Page, D. C. The human Y chromosome: Overlapping DNA clones spanning the euchromatic region. *Science* **258**, 60–66 (1992).
9. Saxena, R. *et al.* Four DAZ genes in two clusters found in AZFc region of human Y chromosome. *Genomics* **67**, 256–267 (2000).
10. Tilford, C. *et al.* A physical map of the human Y chromosome. *Nature* **409**, 943–945 (2001).
11. Shen, P. *et al.* Population genetic implications from sequence variation in four Y chromosome genes. *Proc. Natl Acad. Sci. USA* **97**, 7354–7359 (2000).
12. Sun, C. *et al.* An azoospermic man with a de novo point mutation in the Y-chromosomal gene *USP9Y*. *Nature Genet.* **23**, 429–432 (1999).
13. Kuroda-Kawaguchi, T. *et al.* The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. *Nature Genet.* **29**, 279–286 (2001).
14. Human Genome Sequencing Consortium Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
15. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
16. Schueler, M. G., Higgins, A. W., Rudd, M. K., Gustashaw, K. & Willard, H. F. Genomic and genetic definition of a functional human centromere. *Science* **294**, 109–115 (2001).
17. Tyler-Smith, C. *et al.* Localization of DNA sequences required for human centromere function through an analysis of rearranged Y chromosomes. *Nature Genet.* **5**, 368–375 (1993).
18. Yeh, R. F., Lim, L. P. & Burge, C. B. Computational inference of homologous gene structures in the human genome. *Genome Res.* **11**, 803–816 (2001).
19. Lahn, B. T. & Page, D. C. Functional coherence of the human Y chromosome. *Science* **278**, 675–680 (1997).
20. Page, D. C., Harper, M. E., Love, J. & Botstein, D. Occurrence of a transposition from the X-chromosome long arm to the Y-chromosome short arm during human evolution. *Nature* **311**, 119–123 (1984).
21. Mumm, S., Molini, B., Terrell, J., Srivastava, A. & Schlessinger, D. Evolutionary features of the 4-Mb Xq21.3 XY homology region revealed by a map at 60-kb resolution. *Genome Res.* **7**, 307–314 (1997).
22. Schwartz, A. *et al.* Reconstructing hominid Y evolution: X-homologous block, created by X-Y transposition, was disrupted by Yp inversion through LINE-LINE recombination. *Hum. Mol. Genet.* **7**, 1–11 (1998).

23. Tyler-Smith, C., Taylor, L. & Muller, U. Structure of a hypervariable tandemly repeated DNA sequence on the short arm of the human Y chromosome. *J. Mol. Biol.* **203**, 837–848 (1988).
24. Graves, J. A. & Schmidt, M. M. Mammalian sex chromosomes: Design or accident? *Curr. Opin. Genet. Dev.* **2**, 890–901 (1992).
25. Jegalian, K. & Page, D. C. A proposed path by which genes common to mammalian X and Y chromosomes evolve to become X inactivated. *Nature* **394**, 776–780 (1998).
26. Lahn, B. T. & Page, D. C. Four evolutionary strata on the human X chromosome. *Science* **286**, 964–967 (1999).
27. Delbridge, M. L., Lingenfelter, P. A., Disteche, C. M. & Graves, J. A. M. The candidate spermatogenesis gene *RBMY* has a homologue on the human X chromosome. *Nature Genet.* **22**, 223–224 (1999).
28. Mazeyrat, S., Saut, N., Mattei, M. G. & Mitchell, M. J. *RBMY* evolved on the Y chromosome from a ubiquitously transcribed X-Y identical gene. *Nature Genet.* **22**, 224–226 (1999).
29. Saxena, R. *et al.* The DAZ gene cluster on the human Y chromosome arose from an autosomal gene that was transposed, repeatedly amplified and pruned. *Nature Genet.* **14**, 292–299 (1996).
30. Lahn, B. T. & Page, D. C. Retroposition of autosomal mRNA yielded testis-specific gene family on human Y chromosome. *Nature Genet.* **21**, 429–433 (1999).
31. Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J. & Stahl, F. W. The double-strand-break repair model for recombination. *Cell* **33**, 25–35 (1983).
32. Jackson, J. A. & Fink, G. R. Gene conversion between duplicated genetic elements in yeast. *Nature* **292**, 306–311 (1981).
33. The International SNP Map Working Group. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**, 928–933 (2001).
34. Underhill, P. A. *et al.* Y chromosome sequence variation and the history of human populations. *Nature Genet.* **26**, 358–361 (2000).
35. Adams, M. *et al.* The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195 (2000).
36. Carvalho, A. B., Dobo, B. A., Vrbanc, M. D. & Clark, A. G. Identification of five new genes on the Y chromosome of *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **98**, 13225–13230 (2001).
37. Waterston, R. H. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
38. Bailey, J. A. *et al.* Recent segmental duplications in the human genome. *Science* **297**, 1003–1007 (2002).
39. Stankiewicz, P. & Lupski, J. R. Genome architecture, rearrangements and genomic disorders. *Trends Genet.* **18**, 74–82 (2002).
40. Ferguson-Smith, M. A. Karyotype-phenotype correlations in gonadal dysgenesis and their bearing on the pathogenesis of malformations. *J. Med. Genet.* **2**, 142–155 (1965).
41. Zinn, A. R., Page, D. C. & Fisher, E. M. C. Turner syndrome: The case of the missing sex chromosome. *Trends Genet.* **9**, 90–93 (1993).
42. Rao, E. *et al.* Pseudoautosomal deletions encompassing a novel homeobox gene cause growth failure in idiopathic short stature and Turner syndrome. *Nature Genet.* **16**, 54–63 (1997).
43. Sun, C. *et al.* Deletion of azoospermia factor (AZFa) region of human Y chromosome caused by recombination between HERV15 proviruses. *Hum. Mol. Genet.* **9**, 2291–2296 (2000).
44. Blanco, P. *et al.* Divergent outcomes of intrachromosomal recombination on the human Y chromosome: male infertility and recurrent polymorphism. *J. Med. Genet.* **37**, 752–758 (2000).
45. Kamp, C., Hirschmann, P., Voss, H., Huellen, K. & Vogt, P. H. Two long homologous retroviral sequence blocks in proximal Yq11 cause AZFa microdeletions as a result of intrachromosomal recombination events. *Hum. Mol. Genet.* **9**, 2563–2572 (2000).
46. Repping, S. *et al.* Recombination between palindromes P5 and P1 on the human Y chromosome causes massive deletions and spermatogenic failure. *Am. J. Hum. Genet.* **71**, 906–922 (2002).
47. Carrel, L., Cottle, A. A., Goglin, K. C. & Willard, H. F. A first-generation X-inactivation profile of the human X chromosome. *Proc. Natl Acad. Sci. USA* **96**, 14440–14444 (1999).
48. Osoegawa, K. *et al.* A bacterial artificial chromosome library for sequencing the complete human genome. *Genome Res.* **11**, 483–496 (2001).
49. Zhao, S. *et al.* Human BAC ends quality assessment and sequence analyses. *Genomics* **63**, 321–332 (2000).
50. Watson, J. M., Spencer, J. A., Riggs, A. D. & Graves, J. A. Sex chromosome evolution: Platypus gene mapping suggests that part of the human X chromosome was originally autosomal. *Proc. Natl Acad. Sci. USA* **88**, 11256–11260 (1991).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank R. Giardine, R. Oates and S. Silber for patient samples; and J. Alfoldi, C. Disteche, J. Koubova and J. Lange for comments on the manuscript. This work was supported by the National Institutes of Health and the Howard Hughes Medical Institute.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.C.P. (page_admin@wi.mit.edu). GenBank accession numbers are listed in Fig. 2l and the Supplementary Information.

Hes1 is a target of microRNA-23 during retinoic-acid-induced neuronal differentiation of NT2 cells

Hiroaki Kawasaki & Kazunari Taira

Department of Chemistry and Biotechnology, School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan, and Gene Function Research Center, National Institute of Advanced Industrial Science and Technology (AIST), Central 4, 1-1-1 Higashi, Tsukuba Science City 305-8562, Japan

MicroRNAs (miRNAs) are phylogenetically widespread small RNAs of 18–25 nucleotides in length, and are found in animals and plants. These small RNAs can regulate gene expression at a translational level through interactions with their target messenger RNAs, and they have a role in the development of *Caenorhabditis elegans* and plants. Although more than two hundred miRNAs have been found in mammals, their mRNA targets remain to be identified. Here, we demonstrate that the expression of Hes1, basic helix–loop–helix transcriptional repressor, is regulated by miRNA-23 (miR-23) in NT2 cells. miR-23 is almost complementary to part of the coding region, just upstream of the termination codon, of *Hes1* mRNA. Reduction in the level of miR-23 by small interfering RNAs resulted in the accumulation of Hes1, and hindered the retinoic-acid-induced neuronal differentiation of NT2 cells. Thus, our results indicate that miR-23 regulates the expression of Hes1 at the post-transcriptional level, and participates in retinoic-acid-induced neuronal differentiation of NT2 cells.

Non-coding RNAs including ribosomal RNA, small nuclear RNA, small nucleolar RNA and transfer RNA have roles in a great variety of processes such as chromosome maintenance, gene imprinting, transcriptional regulation, pre-mRNA splicing and the control of mRNA translation¹. One class of non-coding RNAs called miRNAs are small RNAs of 18–25 nucleotides in length that regulate mRNA at a post-transcriptional level^{2–18}. So far, a large number of miRNAs have been discovered in animals and plants^{2,3,6–16}. Among them, *lin-4* and *let-7* have been identified from the genetic analysis of developmental timing in *C. elegans*, and are well characterized^{2–5}. Both *lin-4* and *let-7* act as repressors of their respective target genes, such as *lin-14*, *lin-28* and *lin-41*. Repression by these miRNAs requires the presence of partially complementary sequences in the 3′-untranslated regions (UTRs) of the target mRNAs. Although *lin-14* and *lin-28* are translationally repressed by *lin-4*, these mRNAs can be detected in association with polyribosomes^{19,20}. Thus, *lin-4* regulates expression of the target genes after translational initiation.

miRNAs are first transcribed as a long RNA and then processed to a pre-miRNA of approximately 70 nucleotides²¹. This pre-miRNA is transported to the cytoplasm and processed by the RNase III enzyme dicer to produce mature miRNA^{21–24}. The mature miRNA is incorporated into ribonucleoprotein complexes (miRNPs) including eIF2C2, which functions in RNA interference (RNAi)-mediated gene silencing^{9,16,25}. This miRNA–miRNP complex represses mRNA translation by partially base-pairing to the 3′-UTR of target mRNAs^{2–5,26,27}. However, *Arabidopsis thaliana* miR-171 and miR-165/166 are perfectly complementary to the coding regions of two scarecrow-like (SCL) putative transcription factor family members, *PHAVOLUTA* (PHV) and *PHABULOSA* (PHB) mRNA, respectively^{17,18} except for a few mismatches at the 3′ end of miR-165/166. These miRNAs can induce cleavage of the mRNAs similar to small interfering (si)RNA-mediated mRNA degradation. Thus, miRNAs have two functions; that is, repressing mRNA translation and mediating cleavage of mRNAs. Our study demonstrates that the expression of Hes1 is regulated by miR-23 at the post-transcriptional level during the retinoic-acid-induced neuronal differentiation of NT2 cells.

Hes1 is a target of miR-23 in NT2 cells

To identify the target mRNAs of miR-23, we searched the NCBI DNA database for exact homologies to miR-23. We did not find any target genes that were exact matches to miR-23 but we found a gene for Hes1 that exhibited significant complementarity (77%) to miR-23. Hes1 is a basic helix–loop–helix (bHLH) transcriptional repressor that is expressed in undifferentiated, but not differentiated, cells^{28–30}. It participates in the Notch signalling pathway in mammals and acts as an anti-differentiation factor^{28–30}. The region that was nearly complementary to miR-23 was located in the coding region, near the termination codon, of *Hes1* mRNA (Fig. 1a). Target sites of other miRNAs, such as *let-7*, have been found in 3′-UTRs of target mRNAs^{2–5}, but some miRNAs in plants target the coding regions of specific mRNAs^{17,18}. Moreover, we considered phylogenetic conservation of sequence between human and mouse *Hes1* to be indicative of the target site for miR-23. The target site of mouse *Hes1* mRNA (at nearly the same position as human *HES1*, including the stop codon) exhibited significant complementarity (74%) to mouse miR-23b. A duplex of Hes1 and miR23 was observed using a prediction programme of mRNA secondary structure (MulFold). This observation suggests that the function of miR-23 has been conserved phylogenetically as a regulator of Hes1 in human and mouse.

To confirm whether *HES1* mRNA might be a target of miR-23, we used human NT2 cells: human embryonal carcinoma (EC) cells that differentiate into neural cells on treatment with retinoic acid³¹. We first examined the expression of Hes1 during retinoic-acid-induced differentiation by amplified enzyme-linked immunosorbent assay (ELISA). NT2 cells were treated with retinoic acid (5 μM) for 3 weeks. As shown in Fig. 1b, Hes1 was easily detectable in undifferentiated NT2 cells. By contrast, Hes1 was barely detectable in differentiated NT2 cells. However, as indicated by northern blotting analysis, the level of *HES1* mRNA in both nuclear and cytoplasmic fractions of cells remained unchanged during retinoic-acid-induced differentiation (Fig. 1c). Moreover, in a sucrose sedimentation assay, an association between *HES1* mRNA and polyribosomes was detected in both undifferentiated and differentiated NT2 cells (Supplementary Fig. S1). Similar observations have been reported

in the case of *lin-4* miRNA and its target gene *lin-14* in *C. elegans*^{19,20}. These results suggest that the expression of *Hes1* might be regulated not at the level of cytoplasmic transport of *HES1* mRNA but at the level of translation during differentiation of NT2 cells. Next we examined the level of miR-23 during retinoic-acid-induced differentiation by northern blotting analysis. miR-23 was barely detectable in undifferentiated NT2 cells but was clearly detectable in differentiated NT2 cells (Fig. 1d). These results suggest that expression of miR-23 might be related to differentiation of NT2 cells.

Regulation of expression of the *HES1* gene by miR-23

Next, to examine whether expression of the gene for *Hes1* is regulated by miR-23, we introduced synthetic miR-23 or mutant miR-23 (Fig. 2a) into undifferentiated NT2 cells. When synthetic miR-23 was introduced at 2 μ M into undifferentiated NT2 cells, the intracellular level of *Hes1* fell significantly (Fig. 2b). By contrast, in the presence of synthetic mutant miR-23, the level of *Hes1* in undifferentiated NT2 cells remained unchanged and similar to that in untreated wild-type NT2 cells (Fig. 2b). In addition, the level of *HES1* mRNA remained constant irrespective of the presence or absence of either synthetic miR-23 or mutant miR-23 (Fig. 2c). These results further suggest that synthetic miR-23 might suppress the expression of the gene for *Hes1* at the translational level.

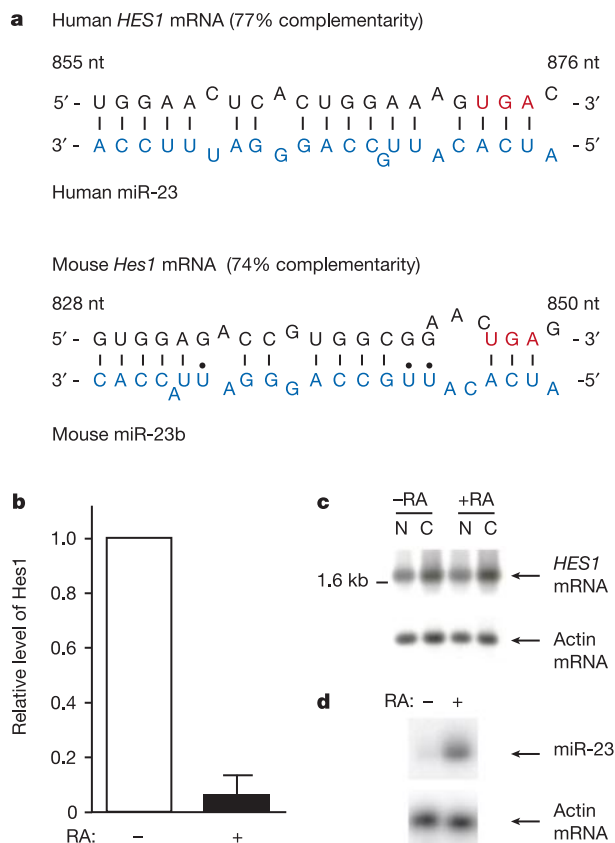


Figure 1 Expression of *Hes1* and miR-23. **a**, Complementarity between *Hes1* mRNA and miR-23. The region nearly complementary to human and mouse miR-23 (blue) is located in the coding region, near the termination codon (red) of human and mouse *Hes1* mRNAs (black). **b**, The level of *Hes1* in NT2 cells in the presence or absence of retinoic acid (RA; 5 μ M, 3 weeks). Values are means (with s.d.) of results from three replicate experiments in each case. **c**, The relative level of *HES1* mRNA in NT2 cells in the presence or absence of retinoic acid (5 μ M, 3 weeks). The relative level of *HES1* mRNA was determined by northern blotting analysis. N, nuclear fraction; C, cytoplasmic fraction. **d**, The level of miR-23 in NT2 cells in the presence or absence of retinoic acid (5 μ M, 3 weeks) determined by northern blotting analysis.

To examine the function of miR-23 further, we tried to reduce the level of endogenous miR-23 using synthetic siRNA (siRNA-miR-23) targeted to a loop region of the precursor to miR-23 (Fig. 2d). siRNAs can induce the RNAi-mediated sequence-specific silencing of gene expression in mammalian cells^{32,33}. RNAi refers to the sequence-specific silencing of gene expression that is induced by double-stranded RNAs in animals and plants^{34,35}. When 100 nM synthetic siRNA-miR-23 was introduced into differentiated NT2 cells, the intracellular level of precursor and mature miR-23 fell significantly (Fig. 2e). By contrast, the level of *Hes1* protein increased in the presence of siRNA-miR-23 (Fig. 2f). Notably,

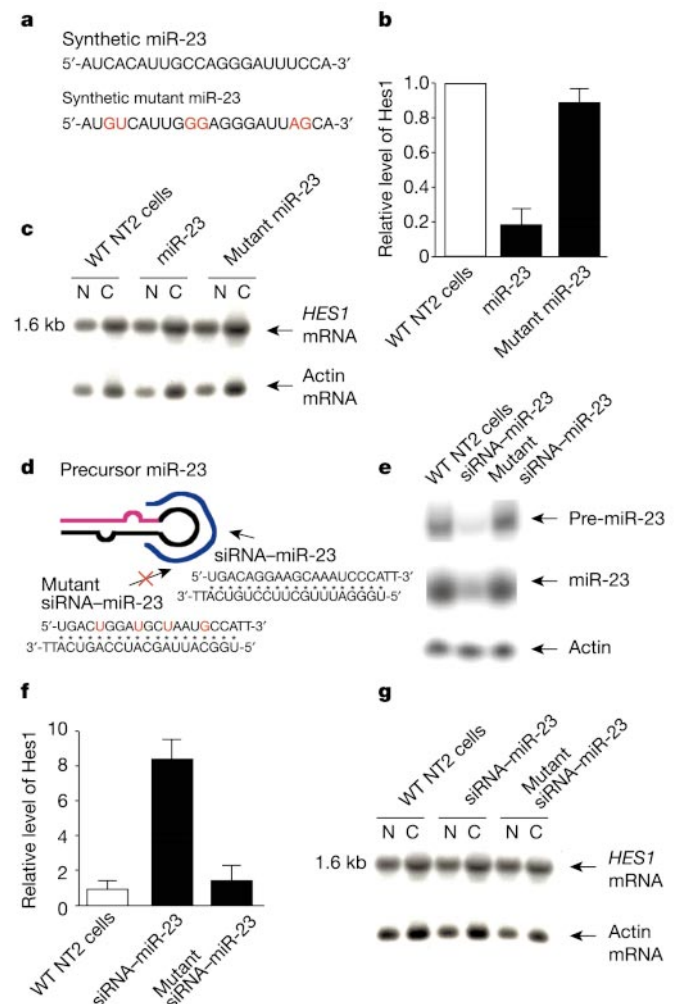


Figure 2 Effects of synthetic miR-23 and siRNA-miR-23 targeted to a loop region of the precursor to miR-23 on expression of the *HES1* gene. **a**, Sequences of synthetic miR-23 and mutant miR-23. Red letters indicate nucleotides different from those in the sequence of miR-23. **b**, The level of *Hes1* in undifferentiated NT2 cells treated with synthetic miR-23 (2 μ M) or with synthetic mutant miR-23 (2 μ M) in the absence of retinoic acid. Values are means (with s.d.) of results from three replicate experiments in each case. WT, wild type. **c**, The level of *HES1* mRNA in undifferentiated NT2 cells treated with synthetic miR-23 or synthetic mutant miR-23 in the absence of retinoic acid. N, nuclear fraction; C, cytoplasmic fraction. **d**, Sequences of synthetic siRNA-miR-23 and synthetic mutant siRNA-miR-23. **e**, The level of precursor and mature miR-23, as detected by northern blotting analysis, in NT2 cells in the presence of retinoic acid (5 μ M, 3 weeks). Actin mRNA was used as an endogenous control. **f**, The level of *Hes1* in differentiated NT2 cells in the presence of retinoic acid (5 μ M). Values are means (with s.d.) of results from three replicate experiments in each case. **g**, The level of *HES1* mRNA in differentiated NT2 cells in the presence of retinoic acid (5 μ M). N, nuclear fraction; C, cytoplasmic fraction.

mutant siRNA-miR-23 (Fig. 2d) did not affect expression of miR-23 or the level of Hes1 in NT2 cells. Moreover, the level of *HES1* mRNA was unaffected by synthetic siRNA-miR-23 (Fig. 2g). These results indicate that synthetic siRNA-miR-23 interfered with the function of miR-23 and, as a result, it promoted the accumulation of Hes1 protein. These results strengthen our hypothesis that miR-23 regulates the expression of Hes1 at the post-transcriptional level.

Target specificity of miR-23 in NT2 cells

To examine the target specificity of miR-23, we constructed plasmids for expression of a chimaeric gene for luciferase that included five copies of the target sequence of miR-23 in *HES1* mRNA (Luc-TS23; Fig. 3a). As a control, we designed a chimaeric gene for luciferase that included the sequence of the mutated target site of miR-23 in *HES1* mRNA (Luc-mTS23; Fig. 3a). Then we introduced each plasmid into NT2 cells and obtained stable cell lines after selection with puromycin. We detected luciferase activity in undifferentiated NT2 cells expressing the gene for Luc-TS23 (Fig. 3b). By contrast, the luciferase activity in differentiated cells expressing Luc-TS23 was lower than that in undifferentiated cells (Fig. 3b). Furthermore, the luciferase activity of Luc-TS23 in undifferentiated NT2 cells that had been treated with synthetic miR-23 was lower than that in untreated wild-type NT2 cells (Fig. 3c). Mutant miR-23 did not affect the luciferase activity of natural TS23-containing Luc-TS23 cells. Notably, a single-stranded mutant miR-23 RNA with appropriate compensatory mutations that restores its pairing to the mTS23 site (but retains the sites of mismatch, as shown at the mutant miR-23 region in Fig. 3a) rescued translational control of Luc-mTS23 (Fig. 3c), highlighting the possibility of creating artificial microRNAs with slightly altered target sequences in the regulation of an arbitrarily chosen target gene. Moreover, the luciferase activity in differentiated NT2 cells that had been treated with synthetic siRNA-miR-23 was higher than that in wild-type

differentiated NT2 cells (Fig. 3d). These results suggest that TS23 in *HES1* mRNA is a target of miR-23 and that the natural near complementarity of TS23 to miR-23 is necessary for miR-23-mediated post-transcriptional silencing of gene expression.

The role of miR-23 during neuronal differentiation

To examine the role of miR-23 during retinoic-acid-induced neuronal differentiation of NT2 cells, we examined a phenotype of NT2 cells grown in the presence or absence of synthetic siRNA-miR-23 by immunostaining with SSEA-3- and MAP2-specific antibodies. SSEA-3 is only expressed in undifferentiated NT2 cells, and MAP2 is only expressed in differentiated NT2 cells^{36,37}. Wild-type NT2 cells differentiate into neural cells on treatment with retinoic acid (Fig. 4a, left panel); however, in the presence of siRNA-miR-23, NT2 cells did not differentiate into neural cells after retinoic acid treatment (Fig. 4a, middle panel). Furthermore, the level of MAP2, a differentiation marker, did not increase after the cells were treated with synthetic siRNA-miR-23, even though the level of MAP2 increased in wild-type differentiated NT2 cells (Fig. 4b). Accordingly, the level of SSEA-3, a marker of undifferentiated cells, did not decrease when NT2 cells were treated with synthetic siRNA-miR-23 and retinoic acid (Fig. 4c). However, the addition of synthetic miR-23 to cells that contained siRNA-miR-23 was able to reverse the effects of siRNA-miR-23, and these cells differentiated into neural cells on treatment with retinoic acid (Fig. 4a, right panel), with an accompanying reduction in the level of SSEA-3 and induction of MAP2 expression. These results suggest that miR-23 has a critical role during retinoic-acid-induced neuronal differentiation.

Discussion

In this study we have identified a target of miRNA in human cells. We demonstrate that *HES1* mRNA is a target of miR-23, and that miR-23 has an important role in retinoic-acid-induced neuronal

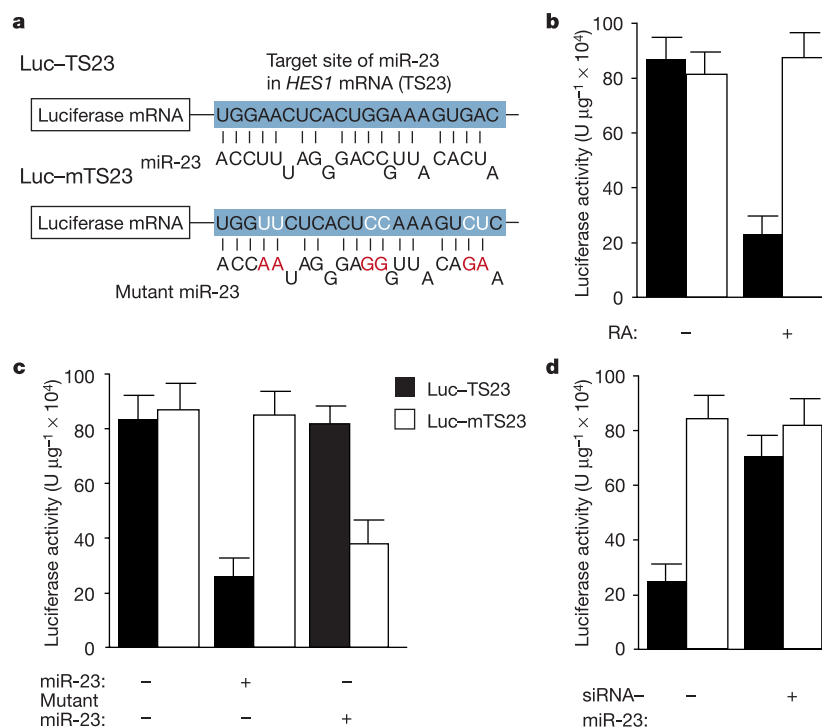


Figure 3 Target specificity of miR-23 as determined with plasmids that encode a gene for luciferase fused to the sequence of the target site of miR-23 in *HES1* mRNA. **a**, Sequences of genes for Luc-TS23 and mutant Luc-TS23 (Luc-mTS23). The blue box indicates the target site of miR-23 or mutant miR-23. White letters indicate nucleotides different from those in the target site of miR-23. **b**, Luciferase activity, owing to the reporter genes, in

NT2 cells in the presence or absence of retinoic acid (5 μM). Values are means (with s.d.) of results from three replicate experiments in each case. **c**, Luciferase activity in undifferentiated NT2 cells in the presence or absence of synthetic miR-23 or mutant miR-23. No retinoic acid was added. **d**, Luciferase activity in differentiated NT2 cells in the presence or absence of siRNA-miR-23. Retinoic acid was added.

differentiation. In general miRNAs, including *lin-4* and *let-7*, control the mRNA translation by partially base-pairing to the 3'-UTR region of their mRNA targets²⁻⁵. Although miR-23 is partially complementary to a coding region—including the termination codon of *HES1* mRNA—it represses the expression of Hes1 at a post-transcriptional level, similar to *lin-4*-mediated repression. Furthermore, in *A. thaliana*, miR-171 and miR-165/166 are perfectly complementary to the coding regions of SCL putative transcription factor, *PHV* and *PHB* mRNA, respectively^{17,18} except for a few mismatches at the 3' end of miR-165/166. These miRNAs cleave their target mRNAs, resulting in siRNA-like gene silencing. It has been proposed that there is only a single pathway shared by both miRNAs and siRNAs, and that this single pathway mediates both translational control and mRNA cleavage³⁸. These findings suggest that there are at least two functional classes of miRNAs; however, the mechanisms of miRNA-mediated gene silencing will need to be studied in detail.

In *C. elegans*, *let-7* and *lin-4* are expressed sequentially during development^{2-5,19}. Thus, as miRNAs suppress the expression of *lin-41* and *lin-14/28* genes, which are necessary for the normal development of *C. elegans*, it is likely that these miRNAs have important roles in development²⁻⁵. In plants, several genes that are targets of miRNAs, including genes in the SCL family, have been identified and their functions have been characterized^{17,18}. The SCL family, which is a target of miR-171, controls a wide range of developmental processes, including radial patterning in roots and hormone signaling. In addition, miR-165/166 can regulate the expression of *PHV* and *PHB* genes that encode homeodomain leucine-zipper transcription factors implicated in the perception of radial position in

the shoot tissues that give rise to leaves. Moreover, *bantam* miRNA simultaneously stimulates cell proliferation and prevents apoptosis during *Drosophila* development³⁹. Thus, these miRNAs are important for the development of animals and plants.

Although more than 200 miRNAs have been found in mammals, the target genes of these miRNAs and their relationship to mammalian development remain to be determined. We identified a target gene (*Hes1*) of a miRNA (miR-23) in mammals, and show that miR-23 participates in retinoic-acid-induced neuronal differentiation of NT2 cells. The *HES1* gene is expressed in undifferentiated cells including neural stem cells, and it inhibits neuronal differentiation, indicating that *HES1* is a negative regulator of differentiation. By contrast, miR-23 is expressed in differentiated NT2 cells and promotes neuronal differentiation through regulation of Hes1 expression at a translational level. Further investigation into the function of miR-23 in neuronal differentiation and development will provide an insight on the role of small non-coding RNAs in neural development. □

Methods

Culture and transfection of cells

Human NT2 cells (Stratagene) were cultured in DMEM medium supplemented with 10% fetal bovine serum (FBS). Transfections were performed with the effectin reagent (Qiagen) according to the manufacturer's protocol. Luc-TS23-expressing and Luc-mTS23-expressing NT2 cells were selected by incubation with puromycin for a week. Retinoic acid was used at 5 μ M for 3 weeks to induce neuronal differentiation of NT2 cells.

Preparation of miRNAs and siRNAs

Synthetic miR-23, mutant miR-23 and siRNAs directed against miR-23 were synthesized with a DNA/RNA synthesizer (model 394; PE Applied Biosystems). For generation of siRNAs, synthetic RNAs were annealed by a standard method³². These siRNAs (100 nM) and synthetic miR-23 (2 μ M) were then introduced into NT2 cells using oligofectamin (Invitrogen) and NeuroPORTER (GTS) according to the manufacturer's protocol.

Construction of plasmids

For construction of the Luc-TS23 and Luc-mTS23 expression plasmids, we used the plasmid pRL-TK (Promega). Five copies of the target site or of the mutant target site of miR-23 were inserted at the *Xba*I site, downstream of the gene for luciferase in pRL-TK. In the case of luciferase reporter genes bearing only one copy of the miR-23 target site, miR-23 barely affected translation of Luc-TS23, probably because of the strong SV40 promoter compared with the natural Hes1 promoter. We confirmed the nucleotide sequence of each chimaeric gene by direct sequencing.

Preparation of nuclear and cytoplasmic fractions of cells

For the preparation of the cytoplasmic fraction, NT2 cells were washed twice with PBS and then resuspended in digitonin lysis buffer (50 mM HEPES/KOH, pH 7.5, 50 mM potassium acetate, 8 mM MgCl₂, 2 mM EGTA and 50 μ g ml⁻¹ digitonin) on ice for 10 min. The lysate was centrifuged at 1,000g and the supernatant was collected as the cytoplasmic fraction. The pellets were resuspended in NP-40 lysis buffer (20 mM Tris-HCl, pH 7.5, 50 mM KCl, 10 mM NaCl, 1 mM EDTA and 0.5% NP-40) and held on ice for 10 min and the resultant lysate was used as the nuclear fraction.

Northern blotting analysis

Cytoplasmic RNA and nuclear RNA were extracted and purified from the cytoplasmic fraction and the nuclear fraction, respectively, with ISOGEN reagent (Wako). Thirty micrograms of total RNA per lane were loaded on a polyacrylamide gel (for detection of miR-23) or agarose gel (for detection of *HES1* mRNA). After electrophoresis, bands of RNA were transferred to a nylon membrane (Amersham). The synthetic DNA probe for Hes1 (5'-GCAGATGCCGACGAGCCTCGAGGCCACC-3') and synthetic RNA probe for miR-23 (5'-GGAAUCCUGGCAUGUGAU-3') were labelled with ³²P by T4 polynucleotide kinase (Takara Shuzo). The level of actin was measured as an endogenous control.

Amplified sandwich ELISA

Amplified ELISA has been described elsewhere⁴⁰. NT2 cells, grown in the presence or absence of retinoic acid (5 μ M, for 3 weeks), were collected. Total protein was used in this assay. ELISA plates were coated with specific polyclonal antibodies against Hes1 (Santa Cruz), SSEA-3 (Santa Cruz) or MAP2 (UBI). After the plates had been washed three times, biotinylated second antibodies, followed by horseradish peroxidase-conjugated streptavidin, were added at room temperature. Absorbance was monitored at 490 nm with a microplate reader after addition of phenylenediamine (Sigma-Aldrich). Normalized levels of Hes1, MAP2 and SSEA-3 were taken arbitrarily as 1.0 (absorbance values at 490 nm for Hes1 in undifferentiated cells, Hes1 in differentiated cells, MAP2 in undifferentiated cells and SSEA-3 in undifferentiated cells were 1.2, 0.06, 0.07 and 0.9, respectively). All values are the average of three experiments with standard deviation.

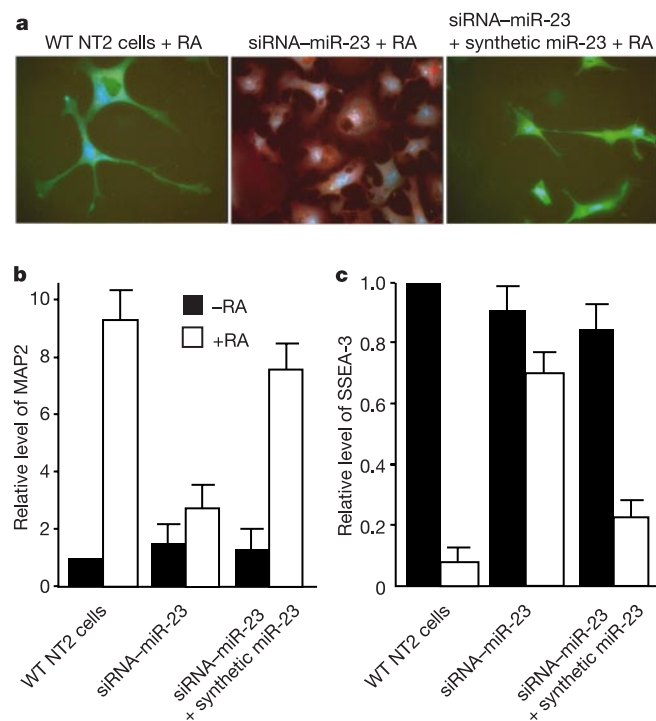


Figure 4 The role of miR-23 during the retinoic-acid-induced neuronal differentiation of NT2 cells **a**, Effects of siRNA-miR-23 on retinoic-acid-induced neuronal differentiation. Left panel, wild-type NT2 cells after treatment with retinoic acid (5 μ M, 3 weeks); middle panel, NT2 cells after treatment with siRNA-miR-23 and retinoic acid; right panel, NT2 cells after treatment with siRNA-miR-23, synthetic miR-23 and retinoic acid. The nuclei of each NT2 cell were stained with DAPI. **b**, The level of MAP2 after retinoic-acid-induced (5 μ M) neuronal differentiation. **c**, The level of SSEA-3 after-retinoic acid-induced (5 μ M) neuronal differentiation.

Assay of luciferase activity

siRNA-miR-23, synthetic miR-23 and mutant miR-23 were introduced into NT2 cells that expressed Luc-TS23 or Luc-mTS23 using oligofectamin (Invitrogen) according to the manufacturer's protocol. After incubation for 72 h, cells were collected and lysed. Total protein was assayed for luciferase activity using a luminometer (Lumat LB9501; Berthold).

Immunostaining

Cells were fixed in paraformaldehyde in PBS for 1 h. Then cells were incubated with polyclonal antibody against SSEA-3 (Santa Cruz) or against MAP2 (UBI) for 2 h. Fluorescein isothiocyanate-conjugated or rhodamine-conjugated secondary antibodies were then added. Nuclei of NT2 cells were stained with 4-diamidino-2-phenylindole (DAPI; Sigma-Aldrich).

Received 9 January; accepted 9 May 2003; doi:10.1038/nature01730.

Published online 8 June 2003.

- Eddy, S. R. Non-coding RNA genes and the modern RNA world. *Nature Rev. Genet.* **2**, 919–929 (2001).
- Lee, R. C., Feinbaum, R. L. & Ambros, V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75**, 843–854 (1993).
- Reinhart, B. J. *et al.* The 21-nucleotide *let-7* RNA regulates developmental timing in *C. elegans*. *Nature* **403**, 901–906 (2000).
- Pasquinelli, A. E. *et al.* Conservation of the sequence and temporal expression of *let-7* heterochronic regulatory RNA. *Nature* **408**, 86–89 (2000).
- Slack, F. J. *et al.* The *lin-41* RBCC gene acts in the *C. elegans* heterochronic pathway between the *let-7* regulatory RNA and the LIN-29 transcription factor. *Mol. Cell* **5**, 659–669 (2000).
- Lau, N. C., Lim, L. P., Weinstein, E. G. & Bartel, D. P. An abundant class of tiny RNAs with probable regulatory roles in *C. elegans*. *Science* **294**, 858–862 (2001).
- Lee, R. C. & Ambros, V. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* **294**, 862–864 (2001).
- Lagos-Quintana, M. *et al.* Identification of novel genes coding for small expressed RNAs. *Science* **294**, 853–858 (2001).
- Mourelatos, Z. *et al.* miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. *Genes Dev.* **16**, 720–728 (2002).
- Reinhart, B. J., Weinstein, E. G., Rhoades, M. W., Bartel, B. & Bartel, D. P. MicroRNAs in plants. *Genes Dev.* **16**, 1616–1626 (2002).
- Lagos-Quintana, M. *et al.* Identification of tissue-specific microRNAs from mouse. *Curr. Biol.* **12**, 735–739 (2002).
- Rhoades, M. W. *et al.* Prediction of plant microRNA targets. *Cell* **110**, 513–520 (2002).
- Lagos-Quintana, M., Rauhut, R., Meyer, J., Borkhardt, A. & Tuschl, T. New microRNAs from mouse and human. *RNA* **9**, 175–179 (2003).
- Lim, L. P., Glasner, M. E., Yekta, S., Burge, C. B. & Bartel, D. P. Vertebrate microRNA genes. *Science* **299**, 1540 (2003).
- Lim, L. P. *et al.* The microRNAs of *Caenorhabditis elegans*. *Genes Dev.* **17**, 991–1008 (2003).
- Dostie, J., Mourelatos, Z., Yang, M., Sharma, A. & Dreyfuss, G. Numerous microRNPs in neuronal cells containing novel microRNAs. *RNA* **9**, 180–186 (2003).
- Llave, C., Xie, Z., Kasschau, K. D. & Carrington, J. C. Cleavage of Scarecrow-like mRNA targets directed by a class of *Arabidopsis* miRNA. *Science* **297**, 2053–2056 (2002).
- Tang, G., Reinhart, B. J., Bartel, D. P. & Zamore, P. D. A biochemical framework for RNA silencing in plants. *Genes Dev.* **17**, 49–63 (2003).
- Olsen, P. H. & Ambros, V. The *lin-4* regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation. *Dev. Biol.* **216**, 671–680 (1999).
- Seggerson, K., Tang, L. & Moss, E. G. Two genetic circuits repress the *Caenorhabditis elegans* heterochronic gene *lin-28* after translation initiation. *Dev. Biol.* **243**, 215–225 (2002).
- Lee, Y., Jeon, K., Lee, J. T., Kim, S. & Kim, V. N. MicroRNA maturation: stepwise processing and

subcellular localization. *EMBO J.* **21**, 4663–4670 (2002).

- Grishok, A. *et al.* Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell* **106**, 23–34 (2001).
- Hutvagner, G. *et al.* A cellular function for the RNA-interference enzyme Dicer in the maturation of the *let-7* small temporal RNA. *Science* **293**, 834–838 (2001).
- Ketting, R. F. *et al.* Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev.* **15**, 2654–2659 (2001).
- Martinez, J., Patkaniowska, A., Urlaub, H., Luhrmann, R. & Tuschl, T. Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell* **110**, 563–574 (2002).
- Zeng, Y., Wagner, E. J. & Cullen, B. R. Both natural and designed micro RNAs can inhibit the expression of cognate mRNAs when expressed in human cells. *Mol. Cell* **9**, 1327–1333 (2002).
- Zeng, Y. & Cullen, B. R. Sequence requirements for microRNA processing and function in human cells. *RNA* **9**, 112–123 (2003).
- Ohtsuka, T. *et al.* Hes1 and Hes5 as notch effectors in mammalian neuronal differentiation. *EMBO J.* **18**, 2196–2207 (1999).
- Kageyama, R., Ohtsuka, T. & Tomita, K. The bHLH gene *Hes1* regulates differentiation of multiple cell types. *Molecules Cells* **29**, 1–7 (2000).
- Ohtsuka, T., Sakamoto, M., Guillemot, F. & Kageyama, R. Roles of the basic helix-loop-helix genes Hes1 and Hes5 in expansion of neural stem cells of the developing brain. *J. Biol. Chem.* **276**, 30467–30474 (2001).
- Cheung, W. M., Chu, P. W., Lung, C. H. & Ip, N. Y. Expression of retinoid receptors during the retinoic acid-induced neuronal differentiation of human embryonal carcinoma cells. *J. Neurochem.* **75**, 34–40 (2000).
- Elbashir, S. M. *et al.* Duplexes of 21-nucleotide RNAs mediate RNA interference in mammalian cell culture. *Nature* **411**, 494–498 (2001).
- McManus, M. T. & Sharp, P. A. Gene silencing in mammals by small interfering RNAs. *Nature Rev. Genet.* **3**, 737–747 (2002).
- Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
- Hammond, S. M., Caudy, A. A. & Hannon, G. J. Post-transcriptional gene silencing by double-stranded RNA. *Nature Rev. Genet.* **2**, 110–119 (2001).
- Pleasant, S. J. & Lee, V. M. NTera 2 cells: a human cell line which displays characteristics expected of a human committed neuronal progenitor cell. *J. Neurosci. Res.* **35**, 585–602 (1993).
- Kurie, J. M. *et al.* Cooperation between retinoic acid and phorbol esters enhances human teratocarcinoma differentiation. *Differentiation* **54**, 115–122 (1993).
- Hutvagner, G. & Zamore, P. D. A microRNA in a multiple-turnover RNAi enzyme complex. *Science* **297**, 2056–2060 (2002).
- Brennecke, J., Hipfner, D. R., Stark, A., Russell, R. B. & Cohen, S. M. Bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene *hid* in *Drosophila*. *Cell* **113**, 25–36 (2003).
- Kawasaki, H. *et al.* Distinct roles of the co-activators p300 and CBP in retinoic-acid-induced F9-cell differentiation. *Nature* **393**, 284–289 (1998).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements The authors thank L. Nelson for critical reading of the manuscript. This research was supported by grants from the Ministry of Economy, Trade and Industry (METI) of Japan, and by a Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Culture (MEXT) of Japan.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.T. (taira@chembio.t.u-tokyo.ac.jp) or H.K. (kawasaki@chembio.t.u-tokyo.ac.jp).

Structure in the early afterglow light curve of the γ -ray burst of 29 March 2003

Makoto Uemura*, Taichi Kato*, Ryoko Ishioka*, Hitoshi Yamaoka†, Berto Monard‡, Daisaku Nogami§, Hiroyuki Maehara||, Atsushi Sugie¶ & Susumu Takahashi¶

* Department of Astronomy, Kyoto University, Kyoto 606-8502, Japan

† Faculty of Science, Kyushu University, Fukuoka 810-8560, Japan

‡ Bronberg Observatory, PO Box 11426, Tiegerpoort 0056, South Africa

§ Hida Observatory, Kyoto University, Kamitakara, Gifu 506-1314, Japan

|| VSOLJ, Namiki 1-13-4, Kawaguchi, Saitama 332-0034, Japan

¶ Dynic Astronomical Observatory, Taga 283-1, Taga-chou, Inugami, Shiga, 522-0341, Japan

Gamma-ray bursts (GRBs) are energetic explosions that for 0.01–100 s are the brightest γ -ray sources in the sky^{1,2}. Observations of the early evolution of afterglows are expected to provide clues about the nature of the bursts, but their rapid fading has hampered such studies; some recent rapid localizations^{3–5} of bursts have improved the situation. Here we report an early detection of the very bright afterglow of the burst of 29 March 2003 (GRB030329). Our data show that, even early in the afterglow phase, the light curve shows unexpectedly complicated structures superimposed on the fading background.

GRB030329 was detected by the HETE-II satellite on 29 March 2003 at 11:37:14.67 UT (ref. 6). A prompt identification of its burst position, which was notified 0.051 days after the burst, enabled observers to make observations during an early phase of the afterglow of the GRB. A new very bright optical source of magnitude 12 was discovered within the error circle of the GRB position, and then continuously monitored by a number of observers^{7–9}. Optical spectroscopic observations determined its redshift to be $z = 0.168$, and using this, its isotropic released energy was calculated to be $E_{\text{iso}} = 8 \times 10^{51} \text{ erg s}^{-1}$ (30–400 keV) (refs 10, 11). In the 30 years of studying GRBs, this GRB afterglow is the closest except for the suspected supernova associated with GRB980425 (ref. 12).

We initiated time-series CCD photometric observations on 29 March 2003 at 12:53:41 UT, about 0.053 days after the burst, at Kyoto University, with 25- and 30-cm telescopes. After the observation at Kyoto, observations at Saitama, the Dynic Astronomical Observatory, and the Bronberg Observatory started, with 20-cm, 60-cm and 30-cm telescopes, respectively. Follow-up observations were performed on 30 March, and 3, 5 and 6 April at Kyoto, on 30 March at Hida, and on 30 and 31 March at Bronberg. A dark-current image was subtracted from obtained CCD images, and then flat-fielding was performed. We calculated the magnitudes of the afterglow with neighbour comparison stars (GSC1434.239, GSC1434.129 and GCS1434.192), whose constancy was better than 0.04 mag during our observation. Our unfiltered CCD observation yields a magnitude system near that of the R_c -system, since the sensitivity peak of the camera is near to the peak of the R_c -system, and the spectra of early afterglows have a smooth continuum without strong emission lines or absorption edges in the optical range. The difference between our unfiltered CCD and the R_c -system is less than 0.02 mag when the spectrum is described with a simple power law ($f(\nu) \propto \nu^p$) with index $-2.0 < p < 0.3$, as expected in GRB optical afterglows¹³.

Our obtained light curve is shown in Fig. 1. In Fig. 1a, the abscissa and the ordinate denote the time from the burst in days and the R_c -magnitude, respectively. To date, the light curve of afterglows, in particular during the early phase of $\lesssim 0.1$ days, has been sampled only sparsely, which has been insufficient to observationally verify

theoretical models. As can be seen in this figure, we succeeded in obtaining a completely continuous light curve from 0.053 to 0.500 days or for 11 hours in GRB030329, which reveals the presence of unambiguous detailed structures in the light curve of early afterglows.

The light curve of the GRB030329 afterglow exhibits clear and repeating deviations from a canonical power-law model. The dotted line in Fig. 1a is the best-fitted power-law model using our light curve and the GRB Coordinates Network data between 0.5 and 10 days. The average decay index, α ($f \propto t^{-\alpha}$), is calculated to be

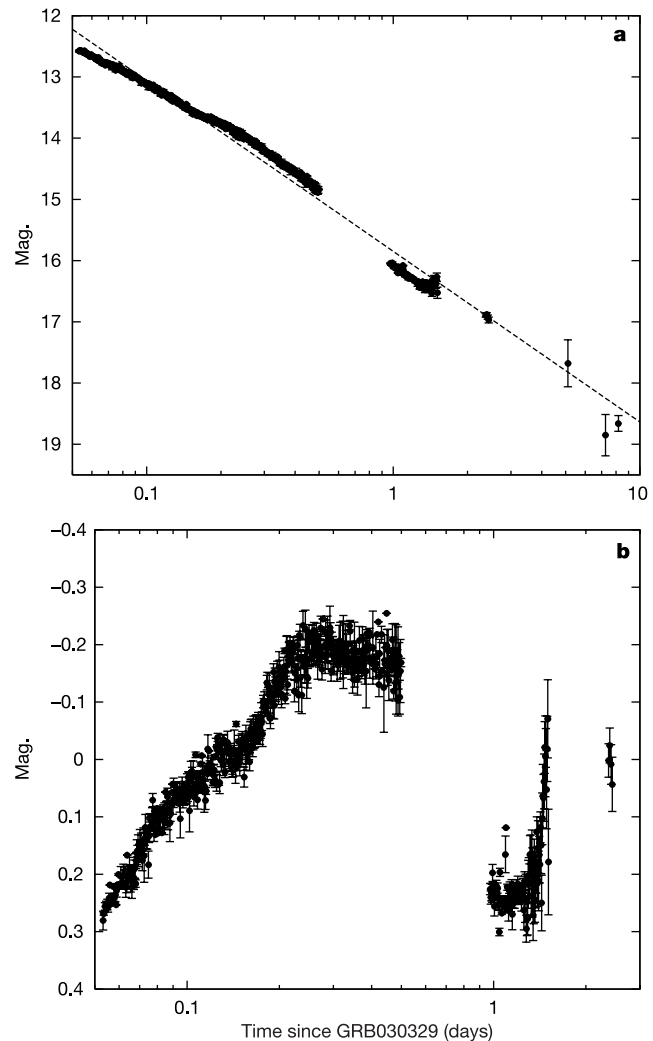


Figure 1 Light curves of the optical afterglow of GRB030329. The abscissa denotes the time from the burst (29 March 2003, 11:37:14.67 UT) in days, shown in logarithmic scale. The ordinate denotes the R_c -magnitude in **a** and the residual magnitude from a power-law decay in **b**. The dotted line in **a** is the best-fitted power-law model, which was used to calculate the residual magnitude in **b**. Our observations are indicated with filled circles with error bars, indicating standard errors. The exposure times of each frame were 30, 40, 60, 30–60 and 45 s for Kyoto, Saitama, the Dynic Astronomical Observatory, the Hida Observatory and the Bronberg observations, respectively. The points in the figure are binned points of the observations with $\Delta \log t_{\text{day}} = 0.002$. Using a smoothly broken power-law model ($f(t) = (f_1(t)^{-n} + f_2(t)^{-n})^{-1/n}$ with $f_i(t) = k_i t^{-\alpha_i}$) (ref. 25), we determined decay indices (α) to be 0.74 ± 0.02 ($0.053 < t < 0.085$), 0.95 ± 0.01 ($0.085 < t < 0.163$), 0.65 ± 0.04 ($0.163 < t < 0.227$), and 1.16 ± 0.01 ($0.227 < t < 0.492$). The break times are 0.085 ± 0.028 , 0.163 ± 0.060 and 0.227 ± 0.043 . Note that the 0.16-day break time was calculated with a broken power-law model without the smoothness parameter. The smoothness parameters, n , are 94 ± 51 for the earlier break and 69 ± 89 for the later break, indicating rapid state transitions.

-1.115 ± 0.006 . This is a standard value for early afterglows of GRBs. As can be seen in the figure, the fading of the afterglow cannot be described only by this simple model. The residual from the single power-law model is shown in Fig. 1b. During the first day, the light curve can be described with four segments, that is, two sets of ascending and descending branches of bumps. Their transitions were not smooth, but occurred in short timespans. A relatively rapid fading presumably follows these bumps. At the end of the first day, the afterglow ceased fading, and then experienced a rebrightening. While they were sampled only partially, the light curve shows that the bumps after 1 day had a rising timescale of 0.1 day. It is interesting to note that an early bump around 0.4 day in Fig. 1 had a timescale of the same order as the later bumps. This similarity might imply that these large (amplitude ~ 0.4 – 0.5 mag) bumps were of the same nature.

The afterglow of GRBs is now widely believed to be synchrotron emission from a forward shock region in an expanding jet colliding with ambient medium, whereas the GRB itself is from an internal shock region between shells^{1,2,14}. In a number of afterglows, it is well known that light curves show a sudden increase of the fading rate due to the jet geometry, typically a few days after the burst^{15,16}. In the case of GRB030329, this jet break is expected to appear later than 60 days from its isotropic energy and redshift^{17,18}. This indicates that the jet break is expected not to appear in Fig. 1, and hence, it is difficult to explain the variation in Fig. 1 with the jet break. As well as the jet break, another well-known phenomenon, the supernova bump, is evidently difficult to use to explain the multiple modulations¹⁹. In GRB030329, the feature of a supernova spectrum appeared 7 days after the burst²⁰. The optical flux before 7 days was hence dominated by the afterglow emission itself.

In some GRB afterglows, similar deviations from a general fading component have been reported^{21–23}. The recent object, GRB021004, is the first object exhibiting multiple fluctuations 1 day after the burst²⁴. In GRB030329, we found that the large-amplitude bumps appeared even in the first day after the burst. The mechanism of these bumps has not been established yet, but the continuous light curve of the bumps of GRB030329 should help us to understand them. □

Received 23 April; accepted 21 May 2003; doi:10.1038/nature01735.

- Wijers, R. A. M. J., Rees, M. J. & Meszaros, P. Shocked by GRB 970228: the afterglow of a cosmological fireball. *Mon. Not. R. Astron. Soc.* **288**, L51–L56 (1997).
- Waxman, E. Gamma-ray-burst afterglow: Supporting the cosmological fireball model, constraining parameters, and making predictions. *Astrophys. J.* **485**, L5–L8 (1997).
- Fox, D. W. *et al.* Early optical emission from the γ -ray burst of 4 October 2002. *Nature* **422**, 284–286 (2003).
- Weidong, L., Alexei, E. V., Ryan, C. & Saurabh, J. The early light curve of the optical afterglow of GRB 021211. *Astrophys. J.* **586**, L9–L12 (2003).
- Fox, D. W. *et al.* Discovery of early optical emission from GRB 021211. *Astrophys. J.* **586**, L5–L8 (2003).
- Vanderspek, R. *et al.* RB030329 (=H2652): A long, extremely bright GRB localized by the HETE WXM and SXC. *GRB Circ. Netw.* 1997 (2003).
- Peterson, B. A. & Price, P. A. GRB 030329: Optical afterglow candidate. *GRB Circ. Netw.* 1985 (2003).
- Torii, K. GRB 030329: OT candidate. *GRB Circ. Netw.* 1986 (2003).
- Uemura, M. GRB 030329: Optical afterglow fading. *GRB Circ. Netw.* 1989 (2003).
- Greiner, J. *et al.* Redshift of GRB 030329. *GRB Circ. Netw.* 2020 (2003).
- Caldwell, N., Garnavich, P., Holland, S., Matheson, T. & Stanek, K. Z. GRB 030329, optical spectroscopy. *GRB Circ. Netw.* 2053 (2003).
- Tinney, C. *et al.* GRB 980425. *IAU Circ.* 6896 (1998).
- Sari, R., Piran, T. & Narayan, R. Spectra and light curves of gamma-ray burst afterglows. *Astrophys. J.* **497**, L17–L20 (1998).
- Sari, R. & Piran, T. Predictions for the very early afterglow and the optical flash. *Astrophys. J.* **520**, 641–649 (1999).
- Rhoads, J. E. The dynamics and light curves of beamed gamma-ray burst afterglows. *Astrophys. J.* **525**, 737–749 (1999).
- Stanek, K. Z., Garnavich, P. M., Kaluzny, J., Pych, W. & Thompson, I. BVRI observations of the optical afterglow of GRB 990510. *Astrophys. J.* **522**, L39–L42 (1999).
- Malesani, D. *et al.* GRB021004: optical observations and predicted break time. *GRB Circ. Netw.* 1607 (2002).
- Frail, D. A. *et al.* Beaming in gamma-ray bursts: Evidence for a standard energy reservoir. *Astrophys. J.* **562**, L55–L58 (2001).
- Reichert, D. E. GRB 970228 revisited: Evidence for a supernova in the light curve and late spectral energy distribution of the afterglow. *Astrophys. J.* **521**, L111–L115 (1999).
- Stanek, K. Z. *et al.* Spectroscopic discovery of the supernova 2003dh associated with GRB 030329. *Astrophys. J.* (in the press); preprint at (<http://xxx.lanl.gov/astro-ph/0304173>) (2003).
- Galama, T. J. *et al.* Optical follow-up of GRB 970508. *Astrophys. J.* **497**, L13–L16 (1998).

- Piro, L. *et al.* Evidence for a late-time outburst of the X-ray afterglow of GB970508 from BeppoSAX. *Astron. Astrophys.* **331**, L41–L44 (1998).
- Garnavich, P. M., Loeb, A. & Stanek, K. Z. Resolving gamma-ray burst 000301C with a gravitational microlens. *Astrophys. J.* **544**, L11–L15 (2000).
- Nakar, E., Piran, T. & Granot, J. Variability in GRB afterglows and GRB 021004. *New Astron.* **8**, 495–505 (2003).
- Beuermann, K. *et al.* VLT observations of GRB 990510 and its environment. *Astron. Astrophys.* **352**, L26–L30 (1999).

Acknowledgements This work is supported partly by a grant-in-aid from the Japanese Ministry of Education, Culture, Sports, Science and Technology. Part of this work is supported by a Research Fellowship of the Japan Society for the Promotion of Science for Young Scientists (M.U. and R.I.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.U. (uemura@kusastro.kyoto-u.ac.jp).

The bright optical afterglow of the nearby γ -ray burst of 29 March 2003

P. A. Price*, D. W. Fox†, S. R. Kulkarni‡, B. A. Peterson*, B. P. Schmidt*, A. M. Soderberg†, S. A. Yost‡, E. Berger†, S. G. Djorgovski†, D. A. Frail§, F. A. Harrison‡, R. Sari||, A. W. Blain† & S. C. Chapman†

* RASA, ANU, Mt Stromlo Observatory, via Cotter Rd, Weston Creek, ACT 2611, Australia

† Caltech Optical Observatories 105-24, ‡ Space Radiation Laboratory 220-47, || Theoretical Astrophysics 130-33, Pasadena, California Institute of Technology, California 91125, USA

§ National Radio Astronomy Observatory, PO Box 0, Socorro, New Mexico 87801, USA

Past studies of cosmological γ -ray bursts (GRBs) have been hampered by their extreme distances, resulting in faint afterglows. A nearby GRB could potentially shed much light on the origin of these events, but GRBs with a redshift $z \leq 0.2$ have been estimated to occur only rarely, about once per decade¹. Here we report the discovery of the bright optical afterglow emission from the burst of 29 March 2003 (GRB030329; ref. 2). The brightness of the afterglow and the prompt report³ of its position resulted in extensive follow-up observations at many wavelengths, along with the measurement of the redshift, $z = 0.169$ (ref. 4). The γ -ray and afterglow properties of GRB030329 are similar to those of GRBs at cosmological redshifts. Observations have already identified the progenitor as a massive star that exploded as a supernova^{5,6}.

On 29 March 2003, 11:37:14.67 UT, GRB030329 triggered all three instruments on board the High Energy Transient Explorer II (HETE-II). About 1.4 hours later, the HETE-II team disseminated via the GRB Coordinates Network (GCN) the 4-arcminute diameter localization² by the Soft X-ray Camera (SXC). We immediately observed the error-circle with the Wide Field Imager (WFI) on the 40-inch telescope at Siding Spring Observatory (SSO) under inclement conditions (nearby thunderstorms). Nevertheless, we were able to identify a bright source not present on the Digitised Sky Survey (Fig. 1) and rapidly communicated the discovery to the community³. The same source was independently detected by the RIKEN automated telescope⁷.

With a magnitude of 12.6 in the *R* band at 1.5 hours, the optical afterglow of GRB030329 is unusually bright. At the same epoch, the well-studied GRB021004 was $R \approx 16$ mag (ref. 8), and the famous GRB990123 was $R \approx 17$ mag (ref. 9). The brightness of this afterglow triggered observations by over 65 optical telescopes around the

world, ranging from submetre apertures to the Keck I 10-metre telescope. Unprecedented bright emission at radio¹⁰, millimetre¹¹, sub-millimetre¹², and X-ray¹³ was also reported (Fig. 2).

Greiner *et al.*⁴ made spectroscopic observations with the Very Large Telescope (VLT) in Chile approximately 16 hours after the GRB and, based on absorption as well as emission lines, announced a redshift of $z = 0.1685$. From Keck spectroscopic observations obtained 8 hours later (Fig. 3) we confirm the VLT redshift, finding $z = 0.169 \pm 0.001$. We note that the optical afterglow of GRB030329 was, at 1.5 hours, approximately the same brightness as the nearest quasar, 3C 273 ($z = 0.158$); it is remarkable that such a large difference in the mass of the engine can produce an optical source with the same luminosity.

With a duration of about 25-s and multi-pulse profile,² GRB030329 is typical of the long duration class of GRBs. The fluence of GRB030329, as detected by the Konus experiment¹⁴, of $1.6 \times 10^{-4} \text{ erg cm}^{-2}$ (in the energy range 15–5,000 keV) places this burst in the top 1% of GRBs.

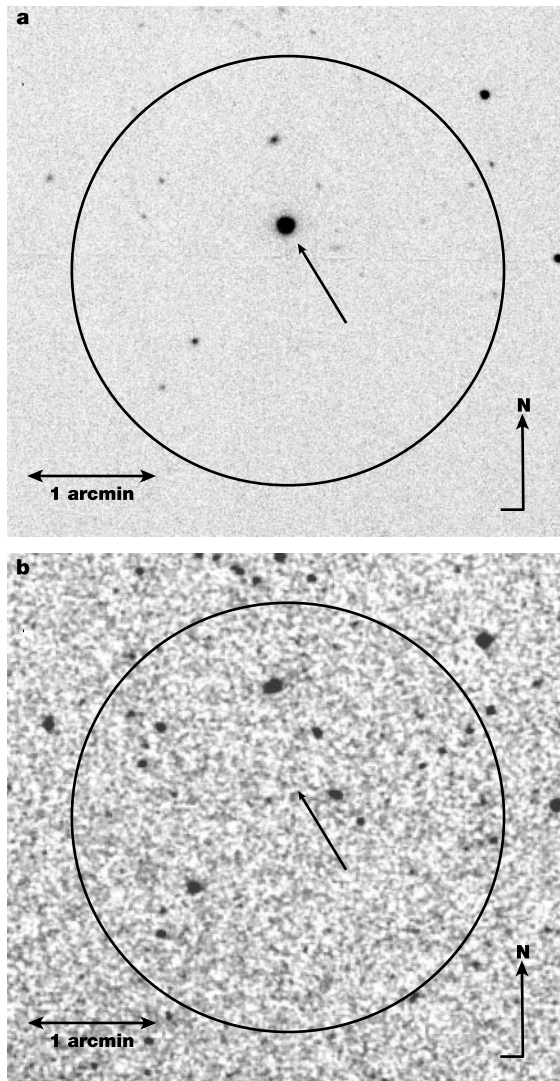


Figure 1 Discovery of the bright optical afterglow of GRB030329. The 600-s exposure taken with the Wide-Field Imager at the 40-inch telescope of the SSO using an *R*-filter (a) started at 13:05 UT, 29 March 2003, about 1.5 hours after the γ -ray event, and was strongly affected by clouds. Nevertheless, comparison of the SSO image with the Second Digitised Sky Survey (b) in the *R_F*-band at the telescope allowed us to identify an afterglow of magnitude 12 (arrowed) within the 4-arcmin HETE-II SXC error-circle² (marked). The position of the optical afterglow was determined with respect to USNO-A2.0 and found to be $\alpha_{2000} = 10:44:59.95$, $\delta_{2000} = +21^{\circ}31'17.8''$ with an uncertainty of 0.5 arcsec in each axis.

At a redshift of 0.169, GRB030329 is the nearest of the cosmological GRBs studied in the 6-year history of afterglow research. Assuming a Lambda cosmology with $H_0 = 71 \text{ km}^{-1} \text{ s}^{-1} \text{ Mpc}^{-1}$, $\Omega_M = 0.27$ and $\Omega_\Lambda = 0.73$, the angular-diameter distance is $d_A = 589 \text{ Mpc}$ and the luminosity distance is $d_L = 805 \text{ Mpc}$. The isotropic γ -ray energy release, $E_{\gamma, \text{iso}} \approx 1.3 \times 10^{52} \text{ erg}$, is typical of cosmological GRBs¹⁵. Likewise, the optical and radio luminosities of the afterglow of GRB030329 are not markedly different from those of cosmological GRBs. In particular, the extrapolated isotropic X-ray luminosity at $t = 10 \text{ h}$ is $L_{X, \text{iso}} \approx 6.4 \times 10^{45} \text{ erg s}^{-1}$, not distinctly different from that of other X-ray afterglows (ref. 16).

Nonetheless, two peculiarities about the afterglow of GRB030329 are worth noting. First, the optical emission steepens from $f(t) \propto t^{-\alpha}$ with $\alpha_1 = 0.873 \pm 0.025$ to $\alpha_2 = 1.97 \pm 0.12$ at epoch $t_* = 0.481 \pm 0.033 \text{ days}$ (Fig. 4; see also ref. 17). This change in α is too large to be due to the passage of a cooling break ($\Delta\alpha = 1/4$; ref. 18) through the optical bands. On the other hand, $\alpha \approx 2$ is typically seen in afterglows following the so-called 'jet-break' epoch (t_j). Before this epoch, the explosion can be regarded as isotropic, and following this epoch the true collimated geometry is manifested. Such an early jet break would imply a substantially-lower energy release than $E_{\gamma, \text{iso}}$.

If $t_* \approx t_j$ then using the formalism and adopting the density and γ -ray efficiency normalizations of ref. 15 we estimate the true γ -ray energy release to be $E_\gamma \approx 3 \times 10^{49} \text{ erg}$. This estimate is 4σ lower than the 'standard energy' of $5 \times 10^{50} \text{ erg}$ found by ref. 15. The geometric-corrected X-ray luminosity¹⁶ would also be the lowest of all X-ray afterglows. If the above interpretation is correct then GRB030329 may be the 'missing link' between cosmological GRBs and the peculiar GRB980425 (ref. 19) ($E_{\gamma, \text{iso}} \approx 10^{48} \text{ erg}$) which has been associated with SN1998bw at $z = 0.0085$.

Second, the decay of the optical afterglow is marked by bumps and wiggles (ref. 20). These features could be due to inhomogeneities in the circumburst medium or additional shells of ejecta from

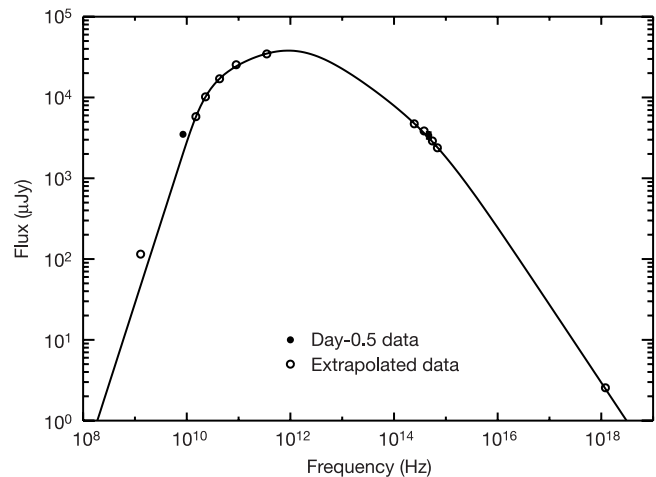


Figure 2 A snapshot spectral flux distribution of the afterglow of GRB030329. This broad-band spectrum of the afterglow at 0.5 days after the GRB^{11–13,25–28} demonstrates the brightness of the afterglow, with the resulting spectral coverage. Solid circles represent measurements made near the nominal time; open circles represent measurements extrapolated to the nominal time assuming an evolution appropriate for a constant-density medium; this figure is therefore meant to be illustrative rather than entirely accurate. A simple fit of an afterglow broad-band spectrum yields the following spectral parameters (we use the convention and symbols of ref. 18): synchrotron self-absorption frequency, $\nu_a \approx 25 \text{ GHz}$; peak frequency, $\nu_m \approx 1270 \text{ GHz}$; cooling frequency, $\nu_c \approx 6.2 \times 10^{14} \text{ Hz}$; peak flux, $f_m \approx 65 \text{ mJy}$; and electron energy index, $p \approx 2$. The physical parameters inferred are as follows: explosion energy, $E \approx 5.7 \times 10^{51} \text{ erg}$; ambient density, $n \approx 5.5 \text{ atom cm}^{-3}$; electron energy fraction, $\epsilon_e \approx 0.16$ and magnetic energy fraction, $\epsilon_B \approx 0.012$.

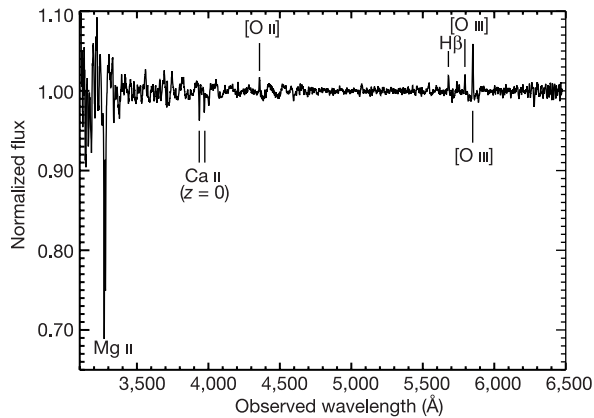


Figure 3 Spectrum of the optical afterglow. The observation was made with the Low Resolution Imaging Spectrometer on the Keck I telescope, using the 400 lines mm^{-1} grism on the blue side, giving an effective resolution of $4.2 \text{ \AA}$. Our observation consisted of a single 600-s exposure on the afterglow. We reduced and extracted the spectra in the standard manner using IRAF. No standard star observations were available, so we have simply fitted and normalized the continuum before smoothing with a $4\text{-\AA}$ boxcar. We identify narrow emission lines from [O II], H β and [O III], and absorption lines from Mg II at a mean redshift, $z = 0.169 \pm 0.001$, making GRB030329 the lowest-redshift cosmological GRB. These emission lines are typical of star-forming galaxies, whereas the absorption lines are caused by gas in the disk of the galaxy. In addition we identify Ca II at $z \approx 0$, presumably due to clouds in our own Galaxy.

the central engine. In either case, the bumps and wiggles complicate the simple jet interpretation offered above. We note that if the GRB had been more distant, and hence the afterglow fainter, then the break in the light curve would probably have been interpreted as the jet break without question.

The proximity of GRB030329 offers us several new opportunities to understand the origin of GRBs. Red bumps in the light curve have been seen in several more-distant ($z \approx 0.3$ to 1) GRB afterglows (for example, refs 21, 22) and interpreted as underlying supernovae that caused the GRBs. While these red bumps appeared to be consistent with a supernova light curve, prior to GRB030329, it had not yet been unambiguously demonstrated that they were indeed supernovae. A clear spectroscopic signature for an underlying supernova has just been identified in the optical afterglow of GRB030329^{5,6}. This demonstrates that the progenitors of at least some GRBs are massive stars that explode as supernovae.

However, there remain a number of issues still to be resolved, relating to the physics of GRB afterglows and the environment around the progenitor. The 'fireball' model of GRB afterglows predicts a broad-band spectrum from centimetre wavelengths to X-rays that evolves as the GRB ejecta expand and sweep up the surrounding medium¹⁸. Testing this model in detail has in the past been difficult, primarily due to both low signal-to-noise ratios and interstellar scintillation at the longer wavelengths, from which come the majority of spectral and temporal coverage of the afterglow evolution (ref. 23). GRB030329, with bright emission at all wavelengths (Fig. 2), and limited scintillation (due to the larger apparent size) will allow astronomers to test the predictions of the fireball model with unprecedented precision through the time evolution of the broad-band spectrum, the angular size of the fireball, and its proper motion (if any). It has long been predicted that if the progenitors of GRBs are massive stars, the circumburst medium should be rich and inhomogeneous²⁴, but it has been difficult to find evidence for this. However, for GRB030329, it should be possible to trace the distribution of circumburst material and determine the environment of the progenitor.

Even in the Swift (launch to be December 2003) era, we expect only one such nearby ($z < 0.2$) GRB every decade (scaling from ref. 1).

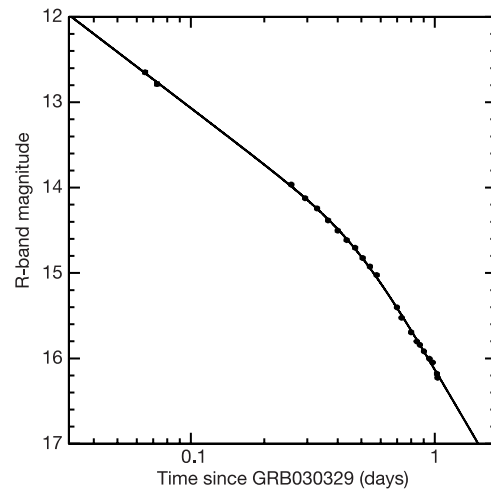


Figure 4 Light curve of the optical afterglow of GRB030329. This *R*-band light curve spans from our discovery to approximately 1 day after the GRB. Owing to the brightness of the GRB, errors in the measurements are smaller than the plotted points. In addition to measurements gleaned from the GCN Circulars^{27,29} we include the following measurements from observations with the SSO 40-inch telescope with WFI in March 2003: 29.5491, $R = 12.649 \pm 0.015$ mag; 29.5568, $R = 12.786 \pm 0.017$ mag; 30.5044, $R = 16.181 \pm 0.010$ mag; 30.5100, $R = 16.227 \pm 0.009$ mag. These measurements are relative to field stars calibrated by Henden³⁰. Solid line, fit of a broken power-law to the data.

Thus GRB030329, the nearest of the cosmological GRBs to date, has given astronomers a rare opportunity to examine a GRB and its afterglow closely. Reports of the many experiments that have been and will be conducted will shed new light on the GRB phenomenon. $\square$

Received 15 April; accepted 21 May 2003; doi:10.1038/nature01734.

- Schmidt, M. Luminosities and space densities of gamma-ray bursts. *Astrophys. J.* **523**, L117–L120 (1999).
- Vanderspek, R., Crew, G., Doty, J., Villaseñor, J., Monnelly, G. *et al.* GRB030329 (=H2652): a long, extremely bright GRB localized by the HETE WXM and SX. *GCN Circ.* 1997 (2003).
- Peterson, B. A. & Price, P. A. GRB 030329: optical afterglow candidate. *GCN Circ.* 1985 (2003).
- Greiner, J. *et al.* Redshift of GRB030329. *GCN Circ.* 2020 (2003).
- Stanek, K. Z. *et al.* Spectroscopic discovery of the supernova 2003dh associated with GRB 030329. *Astrophys. J. Lett.* (in the press); preprint at (<http://arXiv.org/astro-ph/0304173>) (2003).
- Hjorth, J. *et al.* A very energetic supernova associated with the γ -ray burst of 29 March 2003. *Nature* **423**, 847–850 (2003).
- Torii, K. GRB 030329: OT candidate. *GRB Circ. Netw.* 1986 (2003).
- Fox, D. W. *et al.* Early optical emission from the γ -ray burst of 4 October 2002. *Nature* **422**, 284–286 (2003).
- Akerlof, C. *et al.* Observation of contemporaneous optical radiation from a gamma-ray burst. *Nature* **398**, 400–402 (1999).
- Berger, E., Soderberg, A. M. & Frail, D. A. GRB 030329: radio observations. *GCN Circ.* 2014 (2003).
- Kuno, N., Sato, N. & Nakanishi, H. GRB 030329 Radio 23/43/90 GHz observations at Nobeyama. *GCN Circ.* 2089 (2003).
- Hoge, J. C., Meijerink, R., Tilanus, R. P. J. & Smith, I. A. GRB 030329: Sub-millimeter detection. *GCN Circ.* 2088 (2003).
- Marshall, F. & Swank, J. H. RTXE detection of GRB 030329 afterglow. *GCN Circ.* 1996 (2003).
- Golenetskii, S., Mazets, E., Pal'Shin, V., Frederiks, D. & Cline, T. GRB030329a: detection by konus-wind. *GCN Circ.* 2026 (2003).
- Frail, D. A. *et al.* Beaming in gamma-ray bursts: Evidence for a standard energy reservoir. *Astrophys. J.* **562**, L55–L58 (2001).
- Berger, E., Kulkarni, S. R. & Frail, D. A. A standard kinetic energy reservoir in gamma-ray burst afterglows. *Astrophys. J. Lett.* (submitted); preprint at (<http://arXiv.org/astro-ph/0301268>) (2003).
- Garnavich, P., Stanek, K. Z. & Berlind, P. GRB 030329, optical photometry. *GCN Circ.* 2018 (2003).
- Sari, R., Piran, T. & Narayan, R. Spectra and light curves of gamma-ray burst afterglows. *Astrophys. J.* **497**, L17–L20 (1998).
- Galama, T. J. *et al.* An unusual supernova in the error box of the gamma-ray burst of 25 April 1998. *Nature* **395**, 670–672 (1998).
- Lipkin, Y., Ofek, E. O. & Gal-Yam, A. GRB030329—light curve flattening. *GCN Circ.* 2034 (2003).
- Bloom, J. S. *et al.* The unusual afterglow of the gamma-ray burst of 26 March 1998 as evidence for a supernova connection. *Nature* **401**, 453–456 (1999).
- Garnavich, P. M. *et al.* Discovery of the low-redshift optical afterglow of GRB 011121 and its progenitor supernova SN 2001ke. *Astrophys. J.* **582**, 924–932 (2003).
- Berger, E. *et al.* A jet model for the afterglow emission from GRB 000301C. *Astrophys. J.* **545**, 56–62 (2000).
- Chevalier, R. A. & Li, Z.-Y. Gamma-ray burst environments and progenitors. *Astrophys. J.* **520**, L29–L32 (1999).

25. Rao, A. P., Chandra, C. H. I. & Bhattacharya, D. GRB 030329: Radio observations at GMRT. *GCN Circ.* 2073 (2003).
26. Pooley, G. GRB 030329 15 GHz radio observation. *GCN Circ.* 2043 (2003).
27. Burenin, R. *et al.* GRB 030329: light curve observed during the change of its slope. *GCN Circ.* 2024 (2003).
28. Zharikov, S., Benitez, E., Torrealba, J. & Stepanian, J. GRB 030329: SPM optical observations. *GCN Circ.* 2022 (2003).
29. Fitzgerald, J. B. & Orosz, J. A. GRB 030329: optical observations (correction). *GCN Circ.* 2065 (2003).
30. Henden, A. GRB030329. BVRCc field photometry. *GCN Circ.* 2023 (2003).

Acknowledgements P.A.P. and B.P.S. thank the ARC for supporting Australian GRB research. GRB research at Caltech is supported in part by funds from NSF and NASA. We are indebted to S. Barthelmy and the GCN, as well as the *HETE-II* team for prompt alerts of GRB localizations.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.A.P. (pap@mso.anu.edu.au).

A very energetic supernova associated with the γ -ray burst of 29 March 2003

Jens Hjorth*, Jesper Sollerman†, Palle Møller‡, Johan P. U. Fynbo§, Stan E. Woosley||, Chryssa Kouveliotou¶, Nial R. Tanvir#, Jochen Greiner☆☆, Michael I. Andersen**, Alberto J. Castro-Tirado††, José María Castro Cerón‡‡, Andrew S. Fruchter‡‡, Javier Gorosabel‡‡‡, Páll Jakobsson*, Lex Kaper§§, Sylvio Klose|||, Nicola Masetti¶¶, Holger Pedersen*, Kristian Pedersen*, Elena Pian##, Eliana Palazzi¶¶, James E. Rhoads‡‡, Evert Rol§§, Edward P. J. van den Heuvel§§, Paul M. Vreeswijk☆☆, Darach Watson* & Ralph A. M. J. Wijers§§

* Astronomical Observatory, NBI/AFG, University of Copenhagen, Juliane Maries Vej, DK-2100 Copenhagen Ø, Denmark

† Stockholm Observatory, Department of Astronomy, AlbaNova, S-106 91 Stockholm, Sweden

‡ European Southern Observatory, Karl-Schwarzschild-Strasse 2, D-85748 Garching bei München, Germany

§ Department of Physics and Astronomy, University of Aarhus, Ny Munkegade, DK-8000 Århus C, Denmark

|| Department of Astronomy and Astrophysics, University of California, Santa Cruz, California 95064, USA

¶ NSSTC, SD-50, 320 Sparkman Drive, Huntsville, Alabama 35805, USA

Department of Physical Sciences, University of Hertfordshire, College Lane, Hatfield, AL10 9AB, UK

☆☆ Max-Planck-Institut für extraterrestrische Physik, D-85741 Garching, Germany

** Astrophysikalisches Institut, D-14482 Potsdam, Germany

†† Instituto de Astrofísica de Andalucía (IAA-CSIC), PO Box 03004, E-18080 Granada, Spain

‡‡ Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, Maryland 21218, USA

§§ Astronomical Institute Anton Pannekoek, Kruislaan 403, NL-1098 SJ Amsterdam, Netherlands

||| Thüringer Landessternwarte Tautenburg, D-07778 Tautenburg, Germany

¶¶ Istituto di Astrofisica Spaziale e Fisica Cosmica - Sezione di Bologna, CNR, via Gobetti 101, I-40129, Bologna, Italy

INAF, Osservatorio Astronomico di Trieste, Via G.B. Tiepolo 11, I-34131 Trieste, Italy

☆☆ European Southern Observatory, Casilla 19001, Santiago 19, Chile

Over the past five years evidence has mounted that long-duration (>2 s) γ -ray bursts (GRBs)—the most luminous of all astronomical explosions—signal the collapse of massive stars in our Universe. This evidence was originally based on the probable association of one unusual GRB with a supernova¹, but now includes the association of GRBs with regions of massive star

formation in distant galaxies^{2,3}, the appearance of supernova-like ‘bumps’ in the optical afterglow light curves of several bursts^{4–6} and lines of freshly synthesized elements in the spectra of a few X-ray afterglows⁷. These observations support, but do not yet conclusively demonstrate, the idea that long-duration GRBs are associated with the deaths of massive stars, presumably arising from core collapse. Here we report evidence that a very energetic supernova (a hypernova) was temporally and spatially coincident with a GRB at redshift $z = 0.1685$. The timing of the supernova indicates that it exploded within a few days of the GRB, strongly suggesting that core-collapse events can give rise to GRBs, thereby favouring the ‘collapsar’ model^{8,9}.

The first GRB associated with an observed supernova event, SN1998bw, occurred on 25 April 1998 at a redshift of only $z = 0.0085$ (about 37 Mpc)¹. GRB980425 remains by far the closest

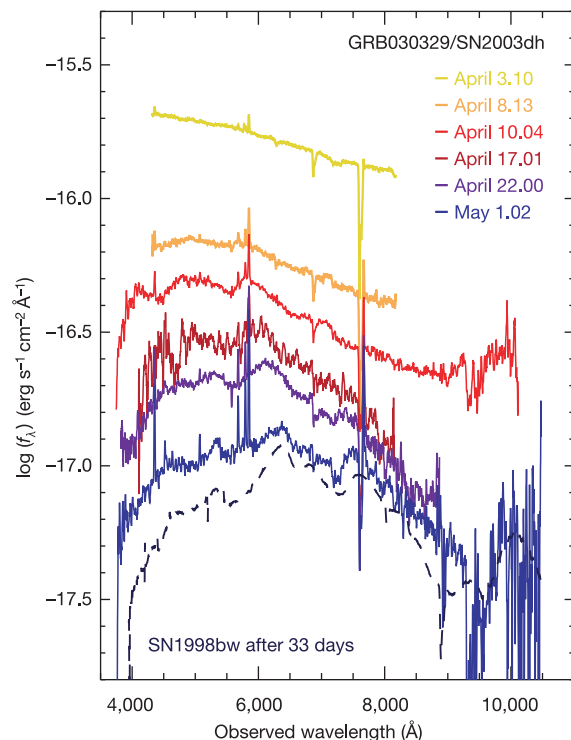


Figure 1 Spectral evolution of the combined optical flux density, f_{λ} , of the afterglow of GRB030329, the associated SN2003dh, and its host galaxy. The details of the observations are given in Table 1. The upper spectrum is rather well fitted by a power law, as usually seen in afterglow spectra. The middle spectra show clear deviations from a power law, similar to SN1998bw at the same phase. The lower spectra, dominated by SN2003dh, reveal the supernova signatures. For comparison, the spectrum of SN1998bw after 33 days (ref. 21) is shown (dashed line) shifted to the GRB030329 redshift. All SN2003dh spectra are presented in observed wavelengths, and no reddening correction has been applied. Telluric absorption lines have also been left in the spectra. The region above 9,000 Å is hampered by sky-line emission, but we tentatively identify a broad feature centred at 10,000 Å that could be due to the supernova. At all epochs we identify emission lines of [O II] λ 3,727, H β , [O III] λ 4,959 and λ 5,007 and H α (Table 2), most probably from the host galaxy. At the last epoch (May 1) we also identify [Ne III] λ 3,869, H δ , H γ , [N II] λ 6,583 as well as blended lines of He I λ 3,889 + H δ and [Ne II] λ 3,968 + He. From the strengths of the Balmer lines we infer that the extinction in the host galaxy is small. The metallicity based on the [O II], [O III], and H β fluxes is [O/H] = -1.0 (ref. 28). The derived star-formation rate is about $0.2 M_{\odot} \text{ yr}^{-1}$ using either [O II] or H α (Table 2). Using the 3σ upper limit of $R > 22.5$ derived for the magnitude of the host from archival data²⁹, we conclude that the equivalent widths of the emission lines are very high. The host galaxy is thus an actively star-forming, low-metallicity, dwarf galaxy and appears to be qualitatively very similar to the host galaxy of GRB980425/SN1998bw (ref. 30).

Table 1 ESO VLT optical spectroscopy of GRB030329/SN2003dh

Date (2003 UT)	Instrument	Grism/Filter	Integration time (s)	Seeing (FWHM in arcseconds)	V magnitude
Apr 3.09	FORS1	V	60	1.10	18.20 ± 0.03
Apr 3.10	FORS1	300V + 10/GG435	1,800	0.6–1.0	20.5 ^{+∞} _{−0.75}
Apr 8.12	FORS1	V	60	0.99	19.30 ± 0.03
Apr 8.13	FORS1	300V + 10/GG435	1,800	0.7–1.3	20.54 ± 0.2
Apr 10.02	FORS2	V	60	1.02	19.65 ± 0.03
Apr 10.04	FORS2	300V + 20/GG375	3,600	0.8–1.8	20.25 ± 0.15
Apr 10.10	FORS2	300I + 21/OG590	5,400	1.4–1.8	
Apr 16.99	FORS2	V	60	1.12	20.16 ± 0.03
Apr 17.01	FORS2	300V + 20/GG375	900	0.67–0.81	20.38 ± 0.2
Apr 21.99	FORS2	V	60	1.01	20.53 ± 0.03
Apr 22.00	FORS2	300V + 20/GG375	900	0.60–1.13	20.95 ± 0.2
Apr 30.99	FORS2	V	60	0.70	21.16 ± 0.05
May 1.01	FORS2	300V + 20/GG375	1,800	0.44–0.68	21.72 ± 0.2
May 1.03	FORS2	300I + 21/OG590	1,800	0.63–0.81	

The epoch of GRB030329 is Mar 29.4842, 2003 UT. The observations were obtained with FORS1 and FORS2 at ESO's 8.2-m Antu and Yepun telescopes at Paranal Observatory, Chile. At all epochs the observations were divided into 600-s or 300-s exposures to facilitate cosmic-ray rejection. A 1.3" wide slit was used, oriented so as to include a nearby star at position angle 123.6°. The data were reduced using IRAF in a standard manner. Flux calibration was performed using a spectrophotometric standard star and absolute flux calibration was done using the simultaneous V-band photometry. On the basis of the spectra of the nearby star we estimate that relative flux calibration errors are less than ±5%. The broad-band V magnitudes are for the combined optical afterglow and supernova. The photometric zero point is based on the secondary standard star USNO U1050_06350247 for which we use $V = 16.84$ (ref. 25). The spectroscopic magnitudes (grism) refer to the derived V magnitudes of the supernova following the decomposition (see Fig. 2). We used the series of fits and decompositions obtained with different templates to estimate the uncertainties in the reported magnitudes and folded in the potential large-scale flux calibration errors. No corrections for foreground reddening ($E(B - V) = 0.025$, ref. 26) were applied.

burst recorded to date. Its proximity and unusually low γ -ray energy budget (total equivalent isotropic energy release of $\sim 8 \times 10^{47}$ erg, about four orders of magnitude less than other long-duration bursts with measured redshifts) have led to some doubt that the progenitor of this GRB was the same as those of GRBs at high redshift. SN1998bw was itself unusual; it displayed kinetic energies more than an order of magnitude greater than typical Ic supernovae and a very high radio flux¹⁰. The tantalizing evidence for the association between GRB980425 and SN1998bw launched a relentless hunt for the 'smoking gun' signature of a supernova in the optical afterglow data of many subsequent GRBs.

On 29 March 2003 (11:37:14.67 UT) NASA's High Energy Transient Explorer, HETE-II, detected a very bright GRB¹¹. GRB030329 lasted about 25 seconds and had a fluence of $\sim 1.2 \times 10^{-4}$ erg cm⁻² and a peak flux of $\sim 7 \times 10^{-6}$ erg cm⁻² s⁻¹ (30–400 keV band). The fluence alone places GRB030329 in the top 0.2% of the 2,704 GRBs detected with the Burst And Transient Source Experiment (BATSE) during its nine years of operation. Rapid follow-up observations^{12,13}, within 1.5 hours, discovered an optical afterglow of magnitude ~ 12 in the R-band, making it brighter than any previously detected afterglow at the same time after burst. A very bright afterglow was also detected at other wavelengths, ranging from X-rays to radio (see ref. 12 and references therein).

Our collaboration¹⁴ determined the redshift of GRB030329, using the Ultraviolet–Visual Echelle Spectrograph (UVES) on the Very Large Telescope (VLT) at the European Southern Observatory (ESO), to be $z = 0.1685$; the second-closest long-duration GRB ever studied, after GRB980425. Assuming a flat $\Omega_{\Lambda} = 0.7$ and $H_0 = 70$ km s⁻¹ Mpc⁻¹ cosmology, GRB030329 was at a luminosity

distance of 810 Mpc. At this distance, GRB030329 is typical in its γ -ray budget: its total isotropic energy release is $\sim 9 \times 10^{51}$ erg (30–400 keV band) and thus qualifies as a classical cosmological GRB. Its relative proximity also accounts for the extreme initial brightness of its optical counterpart¹². Extrapolating the early afterglow power-law decline of the transient and allowing for a supernova light curve similar to that of SN1998bw, we expected to see a supernova emerge 10–20 days after the burst. We therefore conducted spectroscopic observations between 5 and 33 days after the GRB (Table 1), and searched in the spectra of the fading optical afterglow for an emerging supernova contribution.

Figure 1 displays the sequence of six spectra obtained with the VLT Focal Reducer/low dispersion Spectrographs (FORS1 and FORS2) at the epochs listed in Table 1. At the first epoch (3 April), the GRB afterglow spectrum is reasonably well fitted by a single power law with a spectral-index $\beta \approx -1.2 \pm 0.1$ (where $f_{\nu} \propto \nu^{\beta}$), typical for GRB afterglows. By 8 April, it is clear that while the same underlying power-law is present, the small residuals visible on 3 April are now highly significant and resemble the spectrum of a supernova. The absence of broad hydrogen lines indicates a Type I supernova, and the weak or absent Si II $\lambda = 6,150$ Å ($\lambda_{6,150}$) and He lines rule out Types Ia and Ib, respectively. The unusually broad features and blended lines in the spectrum therefore indicate a Type Ic supernova with a very large expansion velocity, remarkably similar to SN1998bw¹⁵. This supernova has been designated SN2003dh following the independent parallel discovery reported in refs 16 and 17.

To monitor the spectral evolution as well as the overall luminosity of SN2003dh, we decomposed the spectra into supernova and

Table 2 Emission-line properties of the host galaxy of GRB030329

ID (2003 UT)	Wavelength (Å)	Redshift	Flux (10 ⁻¹⁷ erg cm ⁻² s ⁻¹)	Luminosity (10 ³⁹ erg s ⁻¹)	SFR (M _⊙ yr ⁻¹)
[O II] λ 3,727	4,355.5 ± 2	0.1685 ± 0.0005	17 ± 2	13 ± 2	0.18 ± 0.06
[Ne III] λ 3,869	4,520.0 ± 2	0.1682 ± 0.0004	3.5 ± 0.7	2.7 ± 0.5	
He I λ 3,889 + H δ	4,542.6 ± 2	0.1681 ± 0.0004	1.1 ± 0.4	0.9 ± 0.3	
[Ne III] λ 3,968 + He	4,638.0 ± 2	0.1686 ± 0.0004	2.4 ± 0.6	1.9 ± 0.5	
H δ λ 4,102	4,792.0 ± 2	0.1683 ± 0.0004	1.1 ± 0.4	0.9 ± 0.3	
H γ λ 4,340	5,071.6 ± 2	0.1684 ± 0.0004	4.5 ± 0.8	3.6 ± 0.6	
H β λ 4,861	5,679.3 ± 2	0.1683 ± 0.0004	10 ± 2	8.1 ± 2	
[O III] λ 4,959	5,793.5 ± 2	0.1683 ± 0.0004	14 ± 2	11 ± 2	
[O III] λ 5,007	5,849.1 ± 2	0.1682 ± 0.0004	40 ± 5	31 ± 4	
H α λ 6,563*	7,667.1 ± 2	0.1683 ± 0.0003	28 ± 3	22 ± 2	0.16 ± 0.05
[N II] λ 6,583	7,693.3 ± 2	0.1686 ± 0.0003	3.3 ± 0.7	2.6 ± 0.5	

The emission line measurements refer to the 1 May spectra (see Table 1). The reported wavelength and redshift uncertainties are dominated by systematic wavelength calibration errors. An accurate estimate of the redshift of the host galaxy was derived using the strong H β and [O III] lines by fixing the wavelength of the nearby [O I] skyline to 5,577.3 Å. The resulting redshift is 0.1686 ± 0.0001 . The properties of the host galaxy of GRB030329 are discussed in Fig. 1. Also shown is the star formation rate (SFR) of the (probably large) fraction of the host galaxy covered by the slit, based on the calibrations given in ref. 27.

*Measurements based on the H α line are marginally affected by telluric lines.

afterglow components, using as templates known spectra of SN1998bw at various stages of its evolution. In Fig. 2 we display the spectra of SN2003dh along with those of SN1998bw at various epochs. The decomposed spectra clearly bring out the emergent supernova features. At 8–10 days (restframe) after the burst, broad lines around 5,000 Å and 5,800 Å (restframe 4,300 Å and 5,000 Å) develop, resembling the SN1998bw spectrum¹⁵. After 28 days new lines emerge at 5,300 Å and 7,500 Å (restframe 4,550 Å and 6,400 Å), presumably due to Mg I λ 4,571 and [O I] λ 6,300 and λ 6,364, characteristic of Type Ic supernova spectra¹⁵.

From the spectra we measure the expansion velocity ten days after GRB030329 from the minimum of the absorption trough of the Si II λ 6,355 line to be $36,000 \pm 3,000 \text{ km s}^{-1}$ ($0.12 \pm 0.01c$). This is

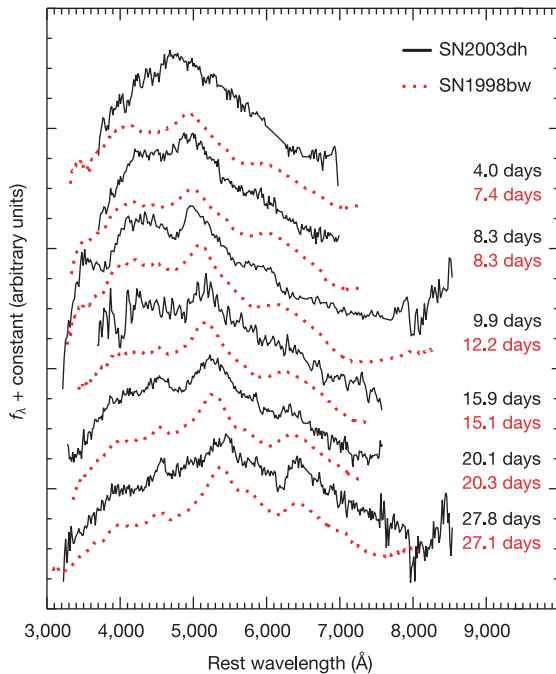


Figure 2 Comparison of the spectral evolution of SN2003dh and SN1998bw. Solid lines indicate spectra of SN2003dh obtained through decomposition as described below. Dotted red lines indicate spectra of SN1998bw taken at similar epochs¹⁵. All spectra are plotted as a function of restframe wavelength and labelled with the time since the GRB, corrected for the cosmological time dilation $(1+z)$; the relative vertical offsets are arbitrary. The SN2003dh spectra are binned to 15-Å bins, the emission lines from the host galaxy have been removed, and where necessary the strong telluric absorption features at 6,800 Å, 7,200–7,400 Å and 7,600 Å have been interpolated. The strong absorption ‘edge’ at 6,150 Å (7,200 Å observed) is probably due to telluric absorption. The spectral decompositions were performed as follows. While the host galaxy has strong emission lines, its continuum flux upper limit is negligible at early epochs and significantly less than the total flux at the later epochs. The contribution from the host galaxy was therefore accounted for by simply removing the emission lines. Model spectra were constructed as a sum of a power law ($f_{\lambda} \propto \lambda^{-(\beta+2)}$) and a scaled version of one of the SN1998bw template spectra¹⁵. For each template, or section thereof, a least-squares fit was obtained through fitting of the three parameters: power-law index β , amplitude of afterglow, and amplitude of supernova. In most cases the best fitting index was found to be $\beta \approx -1.2 \pm 0.05$ which was adopted throughout. We note, however, that the resulting overall spectral shape of the supernova contribution does not depend on the adopted power-law index or template spectrum. The striking similarity between the spectra of these supernovae is clearly seen. The spectral peak wavelength, for both supernovae, is shifting towards the red. The shift is on average $\sim 25 \text{ Å per day}$ for SN2003dh, which is similar to the evolution of the early spectra of SN1998bw. The cause of this shift is the growing opacity in the absorption bluewards of 4,900 Å (rest wavelength). The spectral decompositions provide the fraction of the total flux in the V band that is due to the supernova. The resulting SN2003dh V magnitudes are reported in Table 1 and plotted in Fig. 3.

considerably larger than any previously known supernova (for SN1998bw it was $\sim 23,000 \pm 3,000 \text{ km s}^{-1}$ ten days after the explosion¹⁵). Thus SN2003dh qualifies as the most extreme member of the class of peculiar supernovae colloquially known as ‘hypernovae’: supernovae with very broad lines indicative of expansion velocities in excess of $\sim 30,000 \text{ km s}^{-1}$ at early times¹⁸.

We have attempted to date the supernova explosion of SN2003dh spectroscopically, under the assumption of its parallel spectral evolution with SN1998bw. This assumption introduces a potential systematic uncertainty, but we can qualitatively conclude that SN2003dh began ± 2 days relative to GRB030329 (to this should be added the model-dependent uncertainty of $+0.7/-2$ days of the SN1998bw dating¹⁹). A more detailed analysis will be presented elsewhere (J.S. *et al.*, manuscript in preparation).

As an additional dating check, we have compared the V-band light curves of SN2003dh and SN1998bw (Fig. 3a). SN2003dh peaked about 10–13 days (restframe) after GRB030329 whereas SN1998bw peaked about 17 days after GRB980425. At maximum, SN2003dh was slightly brighter than SN1998bw, consistent with its larger expansion velocity. The difference between the light curves indicates that SN2003dh may have preceded its associated GRB by about 4–7 days, somewhat earlier than in the case of SN1998bw. However,

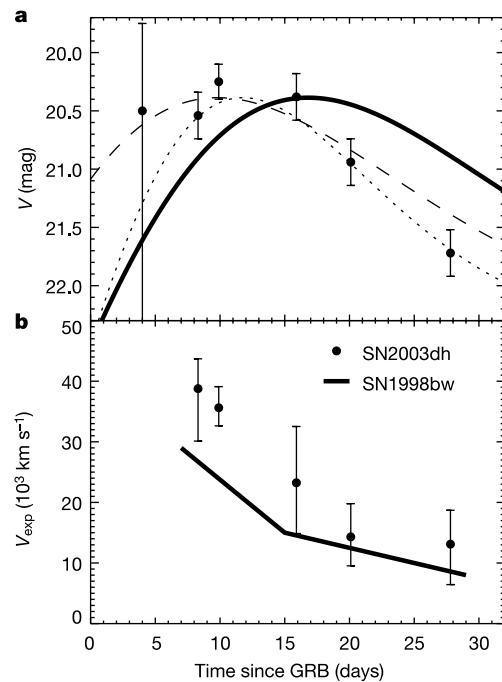


Figure 3 Light curves and expansion velocities of SN2003dh and SN1998bw. **a**, Light curves. Filled circles, spectroscopic V magnitudes of SN2003dh (Table 1) as a function of time (restframe) since GRB030329. Solid line, the brightness of SN1998bw¹ as it would have appeared in the V band at $z = 0.1685$ as a function of time (restframe) since GRB980425. Dashed line, as for the solid line but shifted 7 days earlier. Such an evolution may be expected if the supernova exploded 7 days before the GRB. For SN2003dh, however, this is inconsistent with its spectral evolution (Fig. 2). Dotted line, as for solid line, but evolution speeded up by multiplying time by 0.7. A faster rise and decay may be expected in asymmetric models in which an oblate supernova is seen pole-on. We assumed 0.20 mag (refs 1, 15) extinction for SN1998bw and none for SN2003dh. **b**, Si II λ 6,355 expansion velocities as a function of time (restframe). Filled circles, SN2003dh; solid line, SN1998bw. The SN2003dh values are our best estimates based on the decomposed spectra (Fig. 2). We caution that in some cases these values are very uncertain owing to other features in the spectra around the expected minimum. The solid line shows the trend for SN1998bw based on the data points in ref. 15. The consistent decaying trend in the ejecta velocity of SN2003dh, together with its very high initial value, indicate that we observed very close to the supernova explosion.

given their parallel spectral evolution (Fig. 2) we consider it more likely that the two supernovae were coeval but displayed different light curves, as indicated by the differences in expansion velocity (Fig. 3b), peak brightness, detailed spectral evolution and strength of associated GRB.

We now consider the properties of SN2003dh and SN1998bw in more detail. Both belong to the unusual class of Type Ic supernovae. SN1998bw was well explained by models involving the very energetic explosion of a massive Wolf–Rayet star^{19,20} (a highly evolved star that has lost its outer hydrogen layers to a wind or companion star). Since Type Ic supernovae have light curves powered by radioactive decay, the bright light curve indicates the synthesis of at least several tenths of a solar mass of ⁵⁶Ni in both events (ref. 21 and the current paper). Modelled in spherical symmetry—which is certainly a questionable hypothesis—SN1998bw required $1\text{--}2 \times 10^{52}$ erg of kinetic energy^{19,20}. The energy requirements here are similar. However, it is much more likely that the explosion was asymmetric; in fact the rapid rise and decay of SN2003dh seen in Fig. 3 may indicate that we are viewing an asymmetric supernova along its axis of most rapid expansion^{9,20}. The necessary energy may therefore be considerably smaller²⁰. Further, the larger ratio of afterglow luminosity to supernova luminosity in GRB030329 is also consistent with a burst that has been seen nearly pole-on²², whereas the low luminosity of GRB980425 and its afterglow may have been due to its having been viewed substantially off-axis, or simply because it ejected less highly relativistic matter.

The combined results on SN1998bw and SN2003dh offer the most direct evidence yet that typical, long-duration, energetic GRBs result from the deaths of massive stars. The lack of hydrogen lines in both spectra is consistent with model expectations that the star lost its hydrogen envelope to become a Wolf–Rayet star before exploding. The broad lines are also suggestive of an asymmetric explosion viewed along the axis of most rapid expansion. The large ⁵⁶Ni abundance needed for the light curve cannot have been made by the low-density jet; its mass and solid angle are too small. But it could have been produced in the wind from an accretion disk as matter flowed into a black hole⁹.

The temporal coincidence of SN2003dh with GRB030329 rules out the supranova model²³ for this event. Neither the jet itself, nor any γ -rays it made, could escape until the supernova had expanded for several months. The presence of the large quantities of ⁵⁶Ni necessary to account for the light curve may also be problematic in the magnetar model²⁴ where there is no accretion disk. Producing the ⁵⁶Ni in a shock would require that the pulsar's energy be distributed to a large mass, which is inconsistent with producing a relativistic jet with almost no mass. Our observations thus support the collapsar model^{8,9} for GRBs. □

Received 16 April; accepted 20 May 2003; doi:10.1038/nature01750.

- Galama, T. J. *et al.* An unusual supernova in the error box of the γ -ray burst of 25 April 1998. *Nature* **395**, 670–672 (1998).
- Paczynski, B. Are gamma-ray bursts in star-forming regions? *Astrophys. J.* **494**, L45–L48 (1998).
- Fruchter, A. S. *et al.* Hubble Space Telescope and Palomar imaging of GRB 990123: Implications for the nature of gamma-ray bursts and their hosts. *Astrophys. J.* **519**, L13–L16 (1999).
- Bloom, J. S. *et al.* The unusual afterglow of the γ -ray burst of 26 March 1998 as evidence for a supernova connection. *Nature* **401**, 453–456 (1999).
- Castro-Tirado, A. J. & Gorosabel, J. Optical observations of GRB afterglows: GRB 970508 and GRB 980326 revisited. *Astron. Astrophys. Suppl. Ser.* **138**, 449–450 (1999).
- Garnavich, P. *et al.* Discovery of the low-redshift optical afterglow of GRB 011121 and its progenitor supernova SN 2002ke. *Astrophys. J.* **582**, 924–932 (2003).
- Reeves, J. N. *et al.* The signature of supernova ejecta in the X-ray afterglow of the γ -ray burst 011211. *Nature* **416**, 512–515 (2002).
- Woosley, S. E. Gamma-ray bursts from stellar mass accretion disks around black holes. *Astrophys. J.* **405**, 273–277 (1993).
- MacFadyen, A. I. & Woosley, S. E. Collapsars: Gamma-ray bursts and explosions in 'failed supernovae'. *Astrophys. J.* **524**, 262–289 (1999).
- Kulkarni, S. R. *et al.* Radio emission from the unusual supernova 1998bw and its association with the γ -ray burst of 25 April 1998. *Nature* **395**, 663–669 (1998).
- Vanderspek, R. *et al.* GRB030329 (=H2652): A long, extremely bright GRB localized by the HETE WXM and SXC. *GCN Circ.* 1997 (2003).
- Price, P. A. *et al.* The bright optical afterglow of the nearby γ -ray burst of 29 March 2003. *Nature* **423**, 844–847 (2003).

- Torii, K. GRB 030329: OT candidate *GCN Circ.* 1986 (2003).
- Greiner, J. *et al.* Redshift of GRB 030329. *GCN Circ.* 2020 (2003).
- Patat, F. *et al.* The metamorphosis of SN 1998bw. *Astrophys. J.* **555**, 900–917 (2001).
- Garnavich, P., Matheson, T., Olszewski, E. W., Harding, P. & Stanek, K. Z. GRB 030329 and supernova 2003dh. *IAU Circ.* 8114 (2003).
- Stanek, K. Z. *et al.* Spectroscopic discovery of the supernova 2003dh associated with GRB 030329. *Astrophys. J. Lett.* (in the press); preprint at (<http://arXiv.org/astro-ph/0304173>) (2003).
- Leonard, D. C., Filippenko, A. V., Chornock, R. & Foley, R. J. Photospheric-phase spectropolarimetry and nebular-phase spectroscopy of the peculiar Type Ic supernova 2002ap. *Publ. Astron. Soc. Pacif.* **114**, 1333–1348 (2002).
- Iwamoto, K. *et al.* A hypernova model for the supernova associated with the γ -ray burst of 25 April 1998. *Nature* **395**, 672–674 (1998).
- Woosley, S. E., Eastman, R. G. & Schmidt, B. P. Gamma-ray bursts and Type Ic supernova SN 1998bw. *Astrophys. J.* **516**, 788–796 (1999).
- Sollerman, J. *et al.* SN 1998bw at late phases. *Astrophys. J.* **537**, L127–L130 (2000).
- Granot, J., Panaitescu, A., Kumar, P. & Woosley, S. E. Off-axis afterglow emission from jetted gamma-ray bursts. *Astrophys. J.* **570**, L61–L64 (2002).
- Vietri, M. & Stella, L. Supranova events from spun-up neutron stars: an explosion in search of an observation. *Astrophys. J.* **527**, L43–L46 (1999).
- Wheeler, J., Yi, L., Höflich, P. & Wang, L. Asymmetric supernovae, pulsars, magnetars, and gamma-ray bursts. *Astrophys. J.* **537**, 810–823 (2000).
- Zharikov, S., Benitez, E., Torrealba, J. & Stepanian, J. GRB030329, SPM optical observations. *GCN Circ.* 2022 (2003).
- Schlegel, D. J., Finkbeiner, D. P. & Davis, M. Maps of dust infrared emission for use in estimation of reddening and cosmic microwave background radiation foregrounds. *Astrophys. J.* **500**, 525–553 (1998).
- Kennicutt, R. C. Star formation in galaxies along the Hubble sequence. *Annu. Rev. Astron. Astrophys.* **36**, 189–232 (1998).
- Lee, H., Grebel, E. K. & Hodge, P. W. Nebular abundances of nearby southern dwarf galaxies. *Astron. Astrophys.* **401**, 141–159 (2003).
- Blake, C. & Bloom, J. S. GRB 030329: Optical pre-imaging. *GCN Circ.* 2011 (2003).
- Fynbo, J. U. *et al.* Hubble Space Telescope Space Telescope Imaging Spectrograph imaging of the host galaxy of GRB 980425/SN 1998bw. *Astrophys. J.* **542**, L89–L93 (2000).

Acknowledgements We thank F. Patat for discussions. This paper is based on observations collected by the Gamma-Ray Burst Collaboration at ESO (GRACE) at the European Southern Observatory, Paranal, Chile. We thank the ESO staff at the Paranal Observatory, in particular N. Ageorges, P. Gandhi, S. Hubrig, R. Johnson, C. Ledoux, K. O'Brien, R. Scarpa, T. Szeifert and L. Vanzì, for their help in securing the service mode data reported here. We acknowledge benefits from collaboration within the EU FP5 Research Training Network "Gamma-Ray Bursts: An Enigma and a Tool". This work was also supported by the Danish Natural Science Research Council (SNF). J.P.U.F. and K.P. acknowledge support from the Carlsberg Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.H. (jens@astro.ku.dk).

Beating the superparamagnetic limit with exchange bias

Vassil Skumryev*, Stoyan Stoyanov*, Yong Zhang*, George Hadjipanayis*, Dominique Givord† & Josep Nogués‡

* Department of Physics and Astronomy, University of Delaware, Newark, Delaware 19716, USA

† Laboratoire Louis Néel, CNRS, 38042 Grenoble cedex 9, France

‡ Institut Catalana de Recerca i Estudis Avançats (ICREA) and Departament de Física, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain

Interest in magnetic nanoparticles has increased in the past few years by virtue of their potential for applications in fields such as ultrahigh-density recording and medicine^{1–4}. Most applications rely on the magnetic order of the nanoparticles being stable with time. However, with decreasing particle size the magnetic anisotropy energy per particle responsible for holding the magnetic moment along certain directions becomes comparable to the thermal energy. When this happens, the thermal fluctuations induce random flipping of the magnetic moment with time, and the nanoparticles lose their stable magnetic order and become superparamagnetic⁵. Thus, the demand for further

miniaturization comes into conflict with the superparamagnetism caused by the reduction of the anisotropy energy per particle: this constitutes the so-called 'superparamagnetic limit'^{6,7} in recording media. Here we show that magnetic exchange coupling induced at the interface between ferromagnetic and antiferromagnetic systems^{8,9} can provide an extra source of anisotropy, leading to magnetization stability. We demonstrate this principle for ferromagnetic cobalt nanoparticles of about 4 nm in diameter that are embedded in either a paramagnetic or an antiferromagnetic matrix. Whereas the cobalt cores lose their magnetic moment at 10 K in the first system, they remain ferromagnetic up to about 290 K in the second. This behaviour is ascribed to the specific way ferromagnetic nanoparticles couple to an antiferromagnetic matrix.

The system studied consists of Co nanoparticles embedded either in a paramagnetic matrix (C or Al₂O₃) or in an antiferromagnetic (AFM) matrix (CoO). The samples are grown by sequentially depositing a layer of matrix material of thickness 15–20 nm followed by a layer of nanoparticles, up to a total of 10 repeats (Fig. 1). The Co nanoparticles were produced by gas condensation of sputtered atoms in a so-called 'cluster gun'¹⁰, incorporated inside a conventional sputtering apparatus. Some oxygen (7:1 Ar:O₂ ratio) was allowed inside the sputtering chamber to oxidize the surface of the nanoparticles, forming a Co_{core}CoO_{shell} core-shell structure. This partial oxidation produces a passivating layer¹¹, and, in the samples with an AFM CoO matrix, ensures the formation of a continuous Co-core/CoO-matrix interface. The yield of the cluster gun was controlled to obtain a rather disperse spatial distribution of nanoparticles. The matrix material was deposited either by conventional d.c. sputtering (C) or by reactive sputtering with an O₂ partial pressure of 0.7 mtorr (CoO and Al₂O₃). As revealed by transmission electron microscopy (TEM), the Co particles have a roughly spherical core of $d_{\text{core}} \approx 3\text{--}4\text{ nm}$ in diameter and a face-centred

cubic (f.c.c.) structure (Fig. 1). The shell is f.c.c. CoO with a thickness of about $t_{\text{shell}} \approx 1\text{ nm}$. The Co_{core}CoO_{shell} nanoparticles exhibit a log-normal distribution with a mean $d_{\text{core}} + t_{\text{shell}} \approx 4.7\text{ nm}$ and a standard deviation of 1.1 nm. The average in-plane and out-of plane interparticle distances are around 12 nm, significantly larger than the diameter of the nanoparticles. As the maximum dipolar field, $\mu_0 H_{\text{dip}}$ (where μ_0 is the permeability of vacuum), between two nanoparticles in contact amounts to 0.07 T, in no instance may interparticle dipolar interactions be the source of the effects described below.

The temperature (T) dependence of the magnetic moment $m(T)$ under an applied magnetic field ($\mu_0 H_{\text{app}} = 0.01\text{ T}$) was measured after field cooling (FC) and after zero-field cooling (ZFC). The blocking temperature, T_B , is the temperature at which superparamagnetism sets in. Below T_B , the FC and ZFC magnetization curves are split, whereas above T_B , they coincide as the remanence and coercivity have vanished. For 4-nm particles embedded in an Al₂O₃ (paramagnetic) matrix, $T_B \approx 10\text{ K}$ (Fig. 2). Similar results were obtained for a C matrix. Increasing the particle size to about 7 nm resulted in an enhancement of T_B up to 200 K, in reasonable agreement with published results^{12–19} (which are rather spread due to structural differences affecting the anisotropy K_U , and the effect of dipolar interactions). The hysteresis loops of both systems (with Al₂O₃ and C matrices) measured below T_B are characterized by a small value of coercivity ($\mu_0 H_C = 0.02\text{ T}$) (Fig. 3a) and no loop shift on the magnetic field axis for samples field cooled in fields up to 5 T. Exchange bias⁹ at ferromagnetic (FM)/AFM interfaces is characterized by (1) coercivity enhancement, revealing induced uniaxial or multiaxial anisotropy, and (2) hysteresis loop shift along the field axis after field cooling, revealing unidirectional anisotropy. The lack of exchange bias in the non-magnetic matrix systems can tentatively be attributed to the small thickness of the AFM CoO shell^{11,20}. In any case, the exact interpretation of this behaviour will not affect the discussion presented below.

When such Co_{core}CoO_{shell} particles were deposited without dilution in a paramagnetic matrix and hence in close contact with each other, they were found to exhibit all the features of an exchange biased system (Fig. 3b)—exchange bias field of about $\mu_0 H_{\text{eb}} \approx 0.92\text{ T}$ and enhanced coercivity ($\mu_0 H_C = 0.39\text{ T}$ in the ZFC case and $\mu_0 H_C = 0.59\text{ T}$ in the FC one) at 4.2 K. The bias field decreased as

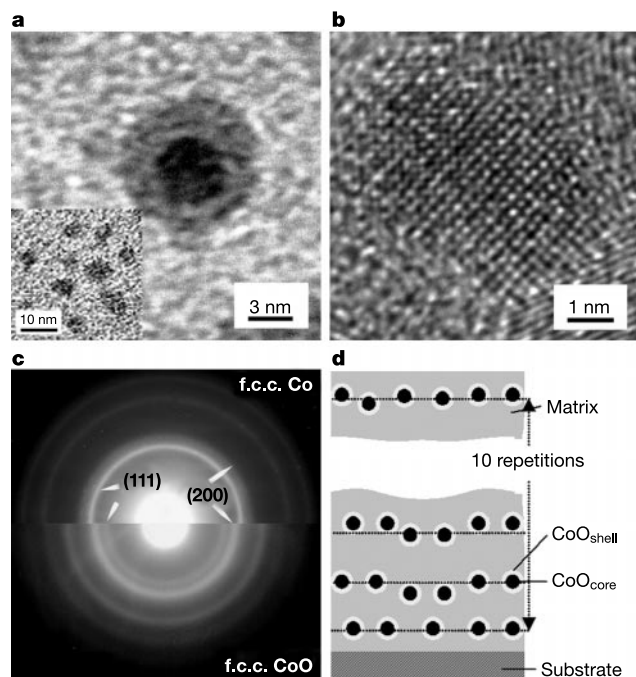


Figure 1 TEM micrographs and electron diffraction of Co_{core}CoO_{shell} particles. **a**, High-magnification bright-field image, revealing the core-shell structure. Inset, plan-view distribution of the nanoparticles. **b**, High-resolution lattice image of nanoparticle with [001]_{f.c.c.} crystallographic orientation. **c**, Electron-diffraction patterns, showing f.c.c. Co and f.c.c. CoO reflections. **d**, Schematic drawing of the sample cross-section, showing Co cores (black), and surrounding CoO shell (white) and matrix (grey).

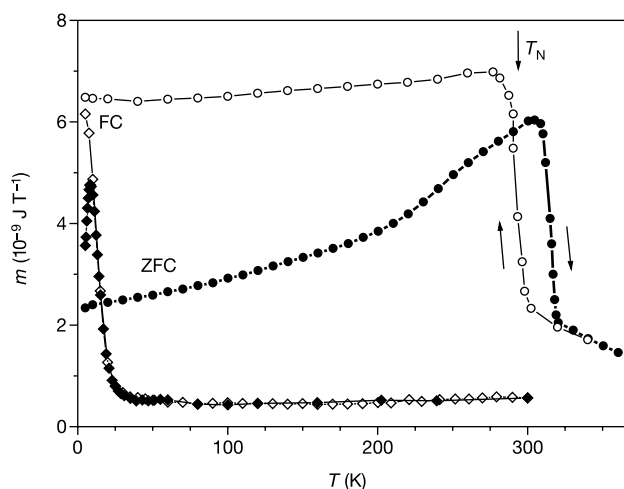


Figure 2 Magnetic moments of 4-nm Co_{core}CoO_{shell} particles. Shown is the temperature dependence of the zero-field cooled (ZFC; filled symbols) and field-cooled (FC; $\mu_0 H_{\text{FC}} = 0.01\text{ T}$, open symbols) magnetic moment (m) of 4-nm Co_{core}CoO_{shell} particles. Particles were embedded in a paramagnetic (Al₂O₃) matrix (diamonds), or in an AFM (CoO) matrix (circles). The measuring field is $\mu_0 H = 0.01\text{ T}$. The Néel temperature of CoO is indicated by an arrow. The lines are guides to the eye.

temperature increased, and vanished at about 180 K, in agreement with the classical data for $\text{Co}_{\text{core}}\text{CoO}_{\text{shell}}$ compacted particles^{8,11,19,21}. This difference in the behaviour of isolated and compacted $\text{Co}_{\text{core}}\text{CoO}_{\text{shell}}$ particles emphasizes the important role of interparticle coupling in stabilizing not only the ferromagnetism of the particle core but also the antiferromagnetism of the particle shell. Note that this stabilization of FM nanoparticle magnetization through interparticle magnetic interactions is not of interest for magnetic recording.

The magnetic behaviour of the isolated $\text{Co}_{\text{core}}\text{CoO}_{\text{shell}}$ particles changes markedly when, instead of being embedded in a paramagnetic matrix, they are embedded in an AFM CoO matrix of similar thickness. As can be seen from the ZFC-FC $m(T)$ curve (Fig. 2), the Co nanocores remain FM up to the Néel temperature, T_N , of CoO ($T_N \approx 290$ K). This indicates that an extra anisotropy is induced such that $K_U V \gg k_B T$, where V is the particle volume and k_B is Boltzmann's constant. In this case, the nanocore moments are prevented from flipping over the energy barrier for all temperatures below T_N of CoO, and thus the nanoparticles remain magnetically stable below this temperature. A hysteresis loop typical of $\text{Co}_{\text{core}}\text{CoO}_{\text{shell}}$ nanoparticles embedded in a CoO matrix is shown in Fig. 3c. Below T_N it is characteristic of an exchange biased system, exhibiting an exchange bias field of $\mu_0 H_{\text{eb}} = 0.74$ T and enhanced coercivity of $\mu_0 H_C = 0.76$ T at 4.2 K. The magnetic stability of the nanoparticles is further evidenced in Fig. 4, where both H_C and the remanent moment, m_R remain much larger than zero for $T < T_N$. In contrast, the Co nanoparticles embedded in a paramagnetic matrix have zero m_R and H_C for $T > T_B = 10$ K. A behaviour similar to the one described above for $\text{Co}_{\text{core}}\text{CoO}_{\text{shell}}$ particles was

also observed for pure Co nanoparticles (that is, without CoO shell) embedded in a CoO matrix. However, because of the poorer quality of the interface between the AFM matrix and the Co FM nanoparticles, all the effects associated with exchange anisotropy were less pronounced. Note that additional measurements on diamagnetic Ag nanoparticles embedded in a CoO matrix did not show FM response, thus confirming that the above effects originate from the FM Co nanocores and not from a possible weak ferromagnetism due to the CoO matrix. Altogether, these results demonstrate that the coupling of FM particles with an AFM matrix can be a source of a large effective anisotropy, leading to a considerable increase of the nanoparticle blocking temperature. We note that the recent observation of loop squareness enhancement in FM particles embedded in or grown on an AFM film^{22,23} has already indicated improved magnetic stability. Moreover, exchange bias has already been proposed theoretically as a possible source of magnetization stability²⁴. However, such large increases in T_B , as well as coercive fields and exchange bias fields of the order of those required to account for the present data, have not been previously observed.

FM-AFM exchange coupling may be understood by assuming that the FM moments of the Co particles create an exchange field, $\mu_0 H_{\text{ex}}$, acting on the interface AFM moments of the CoO matrix. CoO is a collinear antiferromagnet. At a compensated interface, the number of AFM moments from each sublattice is the same, the resulting magnetization is zero and thus the coupling vanishes to first order. To second order, the exchange field induces a canting of the AFM moments, and a FM component is induced along the FM Co moments^{25,26}. Considering that H_{ex} is formally equivalent to an external field, the coupling energy is⁵ $E = -1/2 \chi_{\text{AF}} \mu_0 H_{\text{ex}}^2$, where χ_{AF} is the interface AFM susceptibility. The maximum in χ_{AF} determines the FM 'easy' direction. The energy difference between two FM moment directions, ΔE , amounts to $-(1/2) \Delta \chi_{\text{AF}} \mu_0 H_{\text{ex}}^2$, where $\Delta \chi_{\text{AF}}$ is the associated difference in susceptibility. ΔE may be viewed as an additional anisotropy term, acting as a source of magnetization stability of the Co nanoparticle at finite temperature. Moreover, under an applied field, antiparallel to the FM magnetization, it constitutes a barrier against magnetization reversal. At such very small nanoparticle sizes, the formation of heterogeneous magnetization state costs energy. We found that ΔE is minimum for coherent rotation of the FM moments. The possibility of FM moment rotation within the AFM easy plane, as well as within the plane P perpendicular to the AFM moment direction, were examined. We found that ΔE is minimum for moment rotation within P. The deduced coercive field is $\mu_0 H_C = \Delta E_{\text{min}} / M_S V_{\text{FM}}$, where

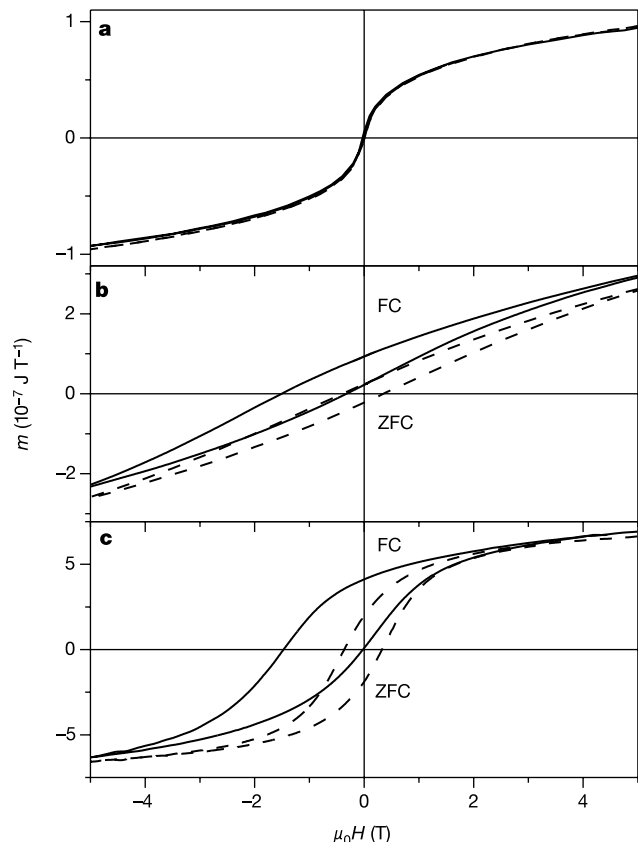


Figure 3 Hysteresis loops at 4.2 K of 4-nm $\text{Co}_{\text{core}}\text{CoO}_{\text{shell}}$ particles embedded in different matrices. Data are shown after ZFC (dashed lines) and FC ($\mu_0 H_{\text{FC}} = 5$ T; solid lines) procedures. **a**, Embedded in a paramagnetic (Al_2O_3) matrix; **b**, compacted; and **c**, embedded in an AFM (CoO) matrix.

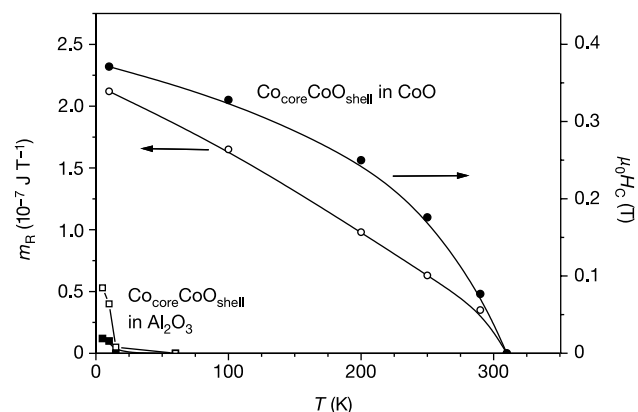


Figure 4 Coercivity and remanence of 4-nm $\text{Co}_{\text{core}}\text{CoO}_{\text{shell}}$ particles. Shown is the temperature dependence of coercivity, H_C (filled symbols), and remanence, m_R (open symbols), after ZFC, of 4-nm $\text{Co}_{\text{core}}\text{CoO}_{\text{shell}}$ particles embedded in different matrices. Squares, a paramagnetic (Al_2O_3) matrix; circles, an AFM (CoO) matrix. The lines are guides to the eye.

$\Delta E_{\min}$ represents the ΔE minimum, V_{FM} is the FM nanoparticle volume and M_S is the saturation magnetization. $\mu_0 H_C = 1.3 \text{ T}$ is calculated for the parameter values of AFM sublattice magnetization⁵ $\mu_0 M' = 1.8 \text{ T}$, a molecular field⁵ coefficient between the AFM sublattices of $w = 100$, an AFM second-order anisotropy constant of $K_1 = 2.7 \times 10^7 \text{ J m}^{-3}$ (ref. 27), $V_{\text{FM}} = 33.5 \text{ nm}^3$ and $\mu_0 H_{\text{ex}} = 84 \text{ T}$ (ref. 28). Considering the uncertainty in the values of most of these parameters, the agreement with the experimental value, $\mu_0 H_C = 0.76 \text{ T}$, is very satisfactory. Unlike the usual exchange-bias systems, the calculated and experimental values of the coercive field are in agreement. Owing to their very small size, the FM particles can be thought of as being coupled to a unique AF domain, and consequently there is no competition between different coupling terms.

Additionally, exchange coupling at the Co/CoO interface gives rise to a shift of the hysteresis loop along the field axis. The large bias fields observed may be shown to suggest the existence of AFM non-collinear moment configurations, reminiscent of spin-glass configurations. Such configurations exist in magnetic oxide nanoparticles^{29,30}, and, by analogy, they may be expected at the surface of small cavities.

In all the above discussion, a compensated Co/CoO interface has been assumed. Coupling energies of the same order of magnitude could be obtained in the case of large uncompensation, of the order of 30%, existing at the Co/CoO interface. However, in this case a significant part of the FM signal would originate from the uncompensated moments in the matrix. The measurements on the Ag/CoO sample do not reveal such signal, and large uncompensation can thus be realistically excluded.

The present results demonstrate that the magnetic coupling of FM nanoparticles with an AFM matrix is a source of a large effective additional anisotropy. This leads to a marked improvement in the thermal stability of the moments of the FM nanoparticles—we observed an increase in the blocking temperature, T_B , of almost two orders of magnitude. This mechanism provides a way to beat the ‘superparamagnetic limit’ in isolated particles. Although it is clear that the system examined here is not suitable in itself for application, the approach developed should in principle apply to nanoparticles deposited on a single AFM layer, a structure suitable for use as a recording medium. With the right choice of FM and AFM components, exchange anisotropy coupling could ultimately allow magnetically stable dots only a few nanometres in size: such dots would surpass the storage-density goal of 1 Tbit in^{-2} , as set by the magnetic-storage industry. □

Received 24 December 2002; accepted 22 April 2003; doi:10.1038/nature01687.

- Kodama, R. H. Magnetic nanoparticles. *J. Magn. Magn. Mater.* **200**, 359–372 (1999).
- Martin, J. I. *et al.* Ordered magnetic nanostructures: Fabrication and properties. *J. Magn. Magn. Mater.* **256**, 449–501 (2003).
- Sun, S. H. *et al.* Monodisperse FePt nanoparticles and ferromagnetic FePt nanocrystal superlattices. *Science* **287**, 1989–1992 (2000).
- Häfele, U., Schütt, W., Teller, J. & Zborowski, M. (eds) *Scientific and Clinical Applications of Magnetic Materials* (Plenum, New York, 1997).
- Chikazumi, S. *Physics of Ferromagnetism* (Oxford Univ. Press, New York, 1997).
- Weller, D. & Moser, A. Thermal effect limits in ultrahigh-density magnetic recording. *IEEE Trans. Magn.* **35**, 4423–4439 (1999).
- Thompson, D. A. & Best, J. S. The future of data storage technology. *IBM J. Res. Dev.* **44**, 311–322 (2000).
- Meiklejohn, W. H. & Bean, C. P. New magnetic anisotropy. *Phys. Rev.* **102**, 1413–1414 (1956).
- Nogués, J. & Schuller, I. K. Exchange bias. *J. Magn. Magn. Mater.* **192**, 203–232 (1999).
- Huberland, H. *et al.* Thin film growth from energetic cluster impact: A feasibility study. *J. Vac. Sci. Technol. A* **10**, 3266–3271 (1992).
- Gangopadhyay, S. *et al.* Exchange anisotropy in oxide passivated Co fine particles. *J. Appl. Phys.* **73**, 6964–6966 (1993).
- Luis, F. *et al.* Magnetic relaxation of interacting Co clusters: Crossover from two- to three-dimensional lattices. *Phys. Rev. Lett.* **88**, 217205 (2002).
- García-Otero, J. *et al.* Influence of dipolar interaction on magnetic properties of ultrafine ferromagnetic particles. *Phys. Rev. Lett.* **84**, 167–170 (2000).
- Gu, E. *et al.* Two-dimensional paramagnetic-ferromagnetic phase transition and magnetic anisotropy in Co(110) epitaxial nanoparticle arrays. *Phys. Rev. B* **60**, 4092–4095 (1999).
- Parent, F. *et al.* Giant magnetoresistance in Co-Ag granular films prepared by low-energy cluster beam deposition. *Phys. Rev. B* **55**, 3683–3687 (1997).

- Held, G. A. *et al.* Competing interactions in dispersions of superparamagnetic nanoparticles. *Phys. Rev. B* **64**, 012408 (2001).
- Woods, S. I. *et al.* Direct investigation of superparamagnetism in Co nanoparticle films. *Phys. Rev. Lett.* **87**, 137205 (2001).
- Murray, C. B. *et al.* Monodisperse 3d transition-metal (Co,Ni,Fe) nanoparticles and their assembly into nanoparticle superlattices. *Mater. Res. Soc. Bull.* **26**, 985–991 (2001).
- Sako, S., Oshima, K. & Saki, M. Magnetic property of oxide passivated Co nanosized particles dispersed in two dimensional plane. *J. Phys. Soc. Jpn* **70**, 2134–2138 (2001).
- Lund, M. S. *et al.* Effect of anisotropy on the critical antiferromagnetic thickness in exchange biased bilayers. *Phys. Rev. B* **66**, 054422 (2002).
- Peng, D. L. *et al.* Magnetic properties of monodispersed Co/CoO clusters. *Phys. Rev. B* **61**, 3103–3109 (2000).
- Sort, J. *et al.* Room temperature coercivity enhancement in mechanically alloyed antiferromagnetic-ferromagnetic powders. *Appl. Phys. Lett.* **75**, 3177–3179 (1999).
- Liu, K. *et al.* Fabrication and thermal stability of arrays of Fe nanodots. *Appl. Phys. Lett.* **81**, 4434–4436 (2002).
- Jensen, P. J. Magnetic recording medium with improved temporal stability. *Appl. Phys. Lett.* **78**, 2190–2192 (2001).
- Schulthess, T. C. & Butler, W. H. Consequences of spin-flop coupling in exchange biased films. *Phys. Rev. Lett.* **81**, 4516–4519 (1998).
- Stamps, R. L. Mechanisms for exchange bias. *J. Phys. D* **33**, R247–R268 (2000).
- Kanamori, J. Theory of the magnetic properties of ferrous and cobaltous oxides. *Prog. Theor. Phys.* **17**, 177–196 (1957).
- Takano, K. *et al.* Interfacial uncompensated antiferromagnetic spins: Role of unidirectional anisotropy in polycrystalline $\text{Ni}_{81}\text{Fe}_{19}/\text{CoO}$ bilayers. *Phys. Rev. Lett.* **79**, 1130–1133 (1997).
- Kodama, R. H., Makhlof, S. H. & Berkowitz, A. E. Finite size effects in antiferromagnetic NiO nanoparticles. *Phys. Rev. Lett.* **79**, 1393–1396 (1997).
- Coe, J. M. D. Noncollinear spin arrangement in ultrafine ferrimagnetic crystallites. *Phys. Rev. Lett.* **27**, 1140–1143 (1971).

Acknowledgements We thank N. Dempsey for critical reading of the manuscript, and D. Weller for discussions. This work was partly supported by the US National Science Foundation, Seagate Technology, the Catalan Direcció General de Recerca, and the Spanish Comisión Interministerial de Ciencia y Tecnología.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to V.S. (vassil@udel.edu).

Sea-level fluctuations during the last glacial cycle

M. Siddall*, E. J. Rohling*, A. Almogi-Labin†, Ch. Hemleben‡, D. Meischner§, I. Schmelzer‡ & D. A. Smeed*

* Southampton Oceanography Centre, European Way, Southampton SO14 3ZH, UK

† Geological Survey of Israel, 30 Malkhe Yisrael Street, Jerusalem 95501, Israel

‡ Department of Geology and Paleontology, University of Tuebingen, Sigwartstrasse 10, D-7400, Tuebingen, Germany

§ Department of Sedimentary Geology, Institute for Geology and Paleontology, University of Göttingen, 37073 Göttingen, Germany

The last glacial cycle was characterized by substantial millennial-scale climate fluctuations^{1–5}, but the extent of any associated changes in global sea level (or, equivalently, ice volume) remains elusive. Highstands of sea level can be reconstructed from dated fossil coral reef terraces^{6,7}, and these data are complemented by a compilation of global sea-level estimates based on deep-sea oxygen isotope ratios at millennial-scale resolution⁸ or higher¹. Records based on oxygen isotopes, however, contain uncertainties in the range of $\pm 30 \text{ m}$, or $\pm 1^\circ \text{C}$ in deep sea temperature^{9,10}. Here we analyse oxygen isotope records from Red Sea sediment cores to reconstruct the history of water residence times in the Red Sea. We then use a hydraulic model of the water exchange between the Red Sea and the world ocean to derive the sill depth—and hence global sea level—over the past 470,000 years (470 kyr). Our reconstruction is accurate to within $\pm 12 \text{ m}$, and gives a centennial-scale resolution from 70 to 25 kyr before present. We

find that sea-level changes of up to 35 m, at rates of up to 2 cm yr^{-1} , occurred, coincident with abrupt changes in climate.

This study takes advantage of the fact that the Red Sea is extremely sensitive to sea-level change, as a consequence of the narrow (18 km) and shallow (137 m) character of its only connection with the open ocean (the Strait of Bab el Mandab). Reduction of the strait profile by sea-level lowering decreases the exchange transport of water masses through the strait. This results in increased residence times of the water within the Red Sea, enhancing the effect of the high rate of evaporation (2.06 m yr^{-1})¹¹ on properties in the Red Sea. The basin thus amplifies the signals of sea-level change, which are recorded in $\delta^{18}\text{O}$ values of foraminifera in Red Sea sediment cores. This amplification has been previously used to calculate sea-level lowstands at times of maximum glaciation during the past 500 kyr (ref. 12), which have since been independently corroborated⁹.

To unlock the potential of Red Sea data for the development of continuous sea-level records (rather than only lowstands) in a manner that is independent of existing methods, we combine a realistic representation of exchange flow in the Strait of Bab el Mandab¹³ with an eddy flux parameterization to represent the Red Sea basin¹⁴. We then calculate salinity and, using routines described previously for the Mediterranean Sea¹⁵, $\delta^{18}\text{O}$ for calcite in equilibrium with ambient water. Changes in the modelled salinity and $\delta^{18}\text{O}$ values are dominated by sea level, and the simulated behaviour of $\delta^{18}\text{O}$ with sea-level change is then applied to translate $\delta^{18}\text{O}$ records from sediment cores into records of past sea-level change. For this purpose, records are selected from relatively shallow (<1,000 m) locations away from the axial trough of the Red Sea, to avoid potential hot-brine-related diagenetic alterations of the foraminiferal calcite.

To ensure that our reconstructions account for any potential change in the basin's climatic forcing through time, we determine a 2σ equivalent confidence limit to our sea-level reconstructions by means of sensitivity tests. These comprise extreme scenarios for the annually integrated evaporation and relative humidity, and an additional temperature uncertainty (see Supplementary Information). Before applying the method to the millennial-scale variability of the last glacial cycle, we discuss several validation

exercises that compare our sea-level results with those from other methods.

We resolve exchange transport through the strait in terms of hydraulic control^{13,16}. Observations¹⁷ and hydraulic modelling¹³ for the present day suggest that exchange of waters across the sill at Hanish is sub-maximal. This is due to the dominant effect of the summer (southwestern) monsoon on the modern Red Sea. It causes upwelling of Gulf of Aden Intermediate Water (GAIW) in the Gulf of Aden, which in turn intrudes into the Red Sea as an intermediate layer¹⁷. The hydraulic model realistically represents this intrusion of GAIW between July and September^{13,16}. At lower sea levels, Hanish sill would be shallower and the GAIW upwelling would have to intensify for the intrusion to occur. Consequently, the sensitivity of the Red Sea to monsoonal variability is reduced at lower sea levels. In a simulation with sea level approximately 120 m lower than today, we find maximal exchange at Hanish sill, with flow over the sill reaching an upper limit determined by the sill geometry and evaporation over the Red Sea. That result corroborates the assumption of maximal exchange at glacial times made in previous studies¹².

The combined 'hydraulic + basin' model conserves buoyancy in the Red Sea so that $Q_3 g' = BLW$, where Q_3 is the volume flux leaving the Red Sea in the lower layer, g' is the reduced gravity, B is the buoyancy flux through the surface of the Red Sea ($3.4 \times 10^{-8} \text{ m}^2 \text{ s}^{-3}$), L is the length of the basin (1,960 km) and W is the width of the basin, which varies with sea level. The buoyancy transport within the basin is limited by a maximum eddy flux¹⁴ $Q_3 = kh_1^2 W g' / fL$, where f is the Coriolis parameter, Q_1 is the flux in the Red Sea in the upper layer, h_1 is the thickness of the upper layer in the Red Sea and k is a constant. Technical details of the model, including an evaluation of the procedures applied, are presented in Supplementary Information.

Although the strait exchange is seasonally resolved, the model basin responds only to the annually integrated sill exchange. Previous hydraulic modelling of the exchange indicates that this is a reasonable approximation¹³, and it allows important basin forcing parameters—evaporation and humidity—to be kept constant throughout the annual cycle. The main run uses evaporation and relative humidity at their modern annually averaged levels of

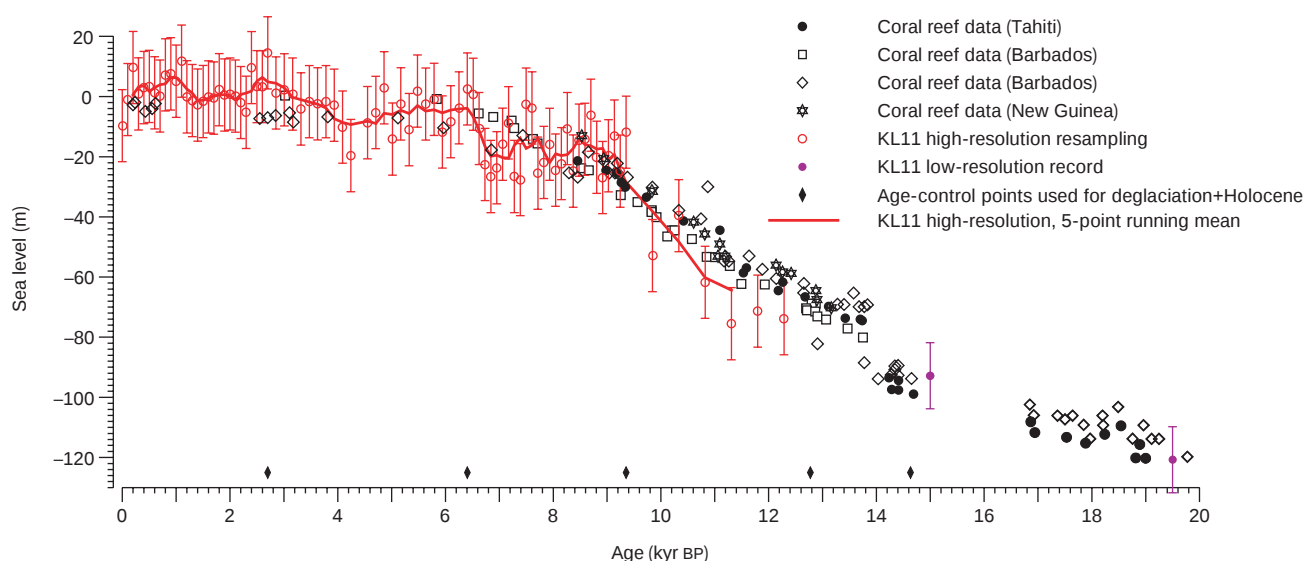


Figure 1 Sea-level reconstruction since the start of the Younger Dryas, based on the $\delta^{18}\text{O}$ record from core KL11 ($18^\circ 44.5' \text{ N}$, $39^\circ 20.6' \text{ E}$) including error bars of $\pm 12 \text{ m}$. Chronology is based on calibrated AMS radiocarbon datings²⁰. The record is zeroed to

modern sea level by removing the mean KL11 record for the past 7 kyr. The point from the low-resolution record of KL11 reported in the LGM is based on intercalibrated benthic $\delta^{18}\text{O}$ data¹⁹. Black symbols are sea-level values obtained from coral reef studies^{21–24}.

2.06 m yr^{-1} (ref. 11) and 70% (ref. 18), respectively. Sea surface temperature is set to decrease linearly with respect to sea level, so that it reaches a value 5 °C lower than today at the Last Glacial Maximum (LGM) (Supplementary Information). Sensitivity tests allow for a ± 2 °C uncertainty in temperature in addition to changes in evaporation between 2.8 m yr^{-1} and 1.4 m yr^{-1} and in relative humidity between 60% and 80%, the modern seasonal extremes^{11,18}. Increases in global ice volume concentrate salinity by ~ 0.01 p.s.u. and $\delta^{18}\text{O}$ by $\sim 0.01\text{‰}$ per metre of sea-level fall^{15,10}. Monsoonal upwelling of GAIW is kept similar to the present throughout the experiments, so that the upper interface of GAIW shallows and deepens to the same extent through the seasonal cycle (that is, sinusoidally from 20 m below the surface in mid-July to 110 m below the surface in mid-January¹³).

Strictly speaking, our sea-level reconstructions pertain to the level at Bab el Mandab, which may deviate somewhat from truly global changes owing to uplift and isostatic effects. Uplift of the strait was previously estimated at $0.044 \pm 0.022 \text{ m kyr}^{-1}$ (ref. 12), and the 470-kyr reconstruction presented here optimizes the agreement with existing data for a rate of 0.02 m kyr^{-1} . Isostatic effects are probably negligible compared to the confidence limits of our method, in view of the landlocked and narrow character of the Red Sea/Gulf of Aden and the great distance from continental ice sheets. Our reconstructions therefore offer close approximations of global sea level (and hence, ice volume), especially during the last glacial cycle, the focus of this study.

We first apply our method to reconstruct sea-level changes since

the onset of the Younger Dryas, based on a $\delta^{18}\text{O}$ series from recent high-resolution resampling of this interval in central Red Sea core GEOTUE-KL11 (referred to here as KL11)^{19,20}, for validation against the sea-level record from coral-reef data^{21–24} (Fig. 1). There is no record from the Red Sea for the LGM and the deglaciation before the onset of the Younger Dryas, because of the presence of a so-called ‘aplanctonic’ interval. Such intervals characterize severe glacial maxima throughout the central and northern Red Sea, and have been related to salinities in excess of 49–50 p.s.u. (refs 12, 19, 25). The sea-level value given for the LGM in KL11 (Fig. 1) derives from intercalibrated $\delta^{18}\text{O}$ measurements of benthic foraminifera¹⁹, and it agrees well with lowstand estimates from the coral reefs. The rest of the sea-level reconstruction from our method also compares favourably with the coral-reef record. Given that the sensitivity of our sea-level method to climatic uncertainties is $\pm 12 \text{ m}$ (equivalent to 2σ confidence limits), the observed deviations of up to 5 m from the mean for individual samples are exactly what should be expected from analyses of a random specimen collection from discrete core samples that integrate decadal, interannual and seasonal climatic variability. Changes in sea level greater than $\pm 12 \text{ m}$ are outside this uncertainty, and hence are considered real.

Next, we compare sea-level reconstructions based on two independent central Red Sea records that both cover several glacial–interglacial cycles, down to 471 kyr before present (BP): cores KL11^{19,20} and MD921017^{12,25} (Fig. 2a). Their close similarity shows the reproducibility of our method (Fig. 2d). This long-core exercise also validates our sea-level reconstructions relative to those from other techniques over the entire glacial–interglacial climate range (Fig. 2c). The good agreement is a clear indication that our $\pm 12 \text{ m}$ confidence limit adequately captures the full range of uncertainty about climatic forcing values. As an additional benefit, our model offers a quantitative expression of the Red Sea salinity history (Fig. 2b).

Having thus validated this method’s capacity to reconstruct sea level with a realistic confidence limit, it is subsequently applied to evaluate any sea-level changes associated with millennial-scale climate fluctuations within the interval 70–25 kyr BP. It was previously established that the influences of the southwestern monsoon did not reach the Red Sea/Gulf of Aden region at those times^{25,26}, effectively ‘locking’ the basin in the present-day winter mode. A discussion of Red Sea insensitivity to monsoon variability is given in Supplementary Information. Hence, the $\pm 12 \text{ m}$ confidence margin to our reconstructions is rather generous as far as this period is concerned.

The 70–25 kyr BP interval in core KL11 was resampled at an average 200-yr resolution, and a $\delta^{18}\text{O}$ record was constructed for *Globigerinoides ruber*²⁰, similar to the previously mentioned long-core and deglaciation records of KL11 and MD921017. An initial rough chronology was established from isotope stratigraphy in the long-core record of KL11¹⁹ (Fig. 2) and from accelerator mass spectrometry (AMS) ^{14}C datings, calibrated²⁷ using a 400-yr reservoir age²⁰ (Fig. 3a). A recent high-resolution deep-sea $\delta^{18}\text{O}$ study¹ suggests that millennial-scale sea-level variability during the last glacial cycle should resemble, in both timing and structure, the Antarctic temperature history recorded in ice-core $\delta^{18}\text{O}$ and δD data^{2,3}. To refine the initial chronology for the 70–25 kyr BP interval in KL11, therefore, we focused on synchronization with the high-resolution Antarctic records (Fig. 3b). There is a remarkable signal similarity (Fig. 3d, e). To refine the initial chronology for the 70–25 kyr BP interval in KL11, therefore, we focused on synchronization with the high-resolution Antarctic records (Fig. 3b, c). There is a remarkable signal similarity (Fig. 3d, e).

Application of our sea-level method to the highly resolved 70–25 kyr BP $\delta^{18}\text{O}$ series of central Red Sea core KL11 yields a record of sea-level variability that agrees well with fossil-reef data^{6,7} (Fig. 4a). We also compare our result with a scaled plot of the high-resolution deep-sea $\delta^{18}\text{O}$ record of eastern North Atlantic core MD952042¹,

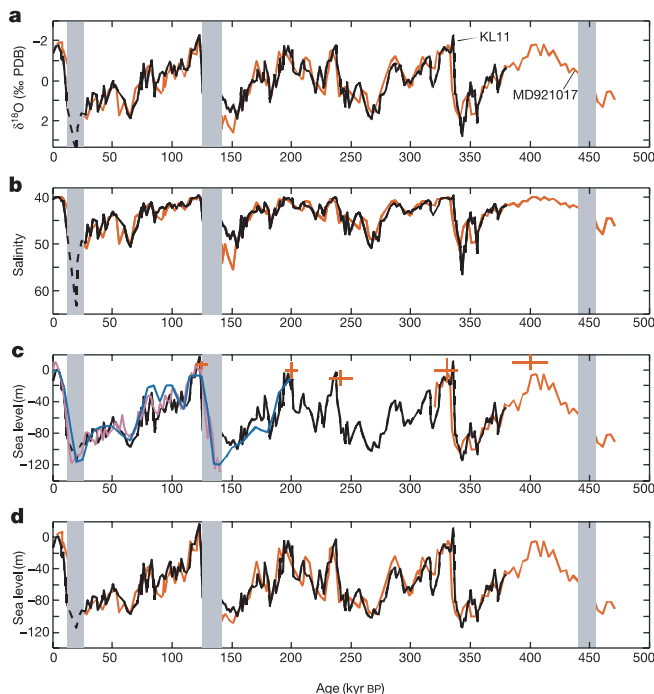


Figure 2 Sea-level reconstruction for the Red Sea for the past 470 kyr. **a**, The long-timescale oxygen isotope records of cores KL11¹⁹ (black line) and MD921017 (19° 23.24' N, 38° 40.84' E)^{12,25} (red line). Grey bands indicate aplanctonic periods in the record. An LGM value based on intercalibrated $\delta^{18}\text{O}$ data¹⁹ is indicated with a dashed line. **b**, The model-generated salinity record from core KL11 (black line) extended with core MD921017 (red line). Salinity in p.s.u. (practical salinity units). **c**, The modelled long-term sea-level record from core KL11 (black line) extended with the record from core MD921017^{12,25} (red line). The pink and blue lines are given for comparison, and represent sea-level compilations from deep-sea oxygen isotope and coral-reef data^{8,22}. Red crosses and bars represent a summary of estimates from several studies^{7,9,12,28,29}. **d**, Comparison of the modelled sea levels from cores KL11 (black line) and MD921017 (red line).

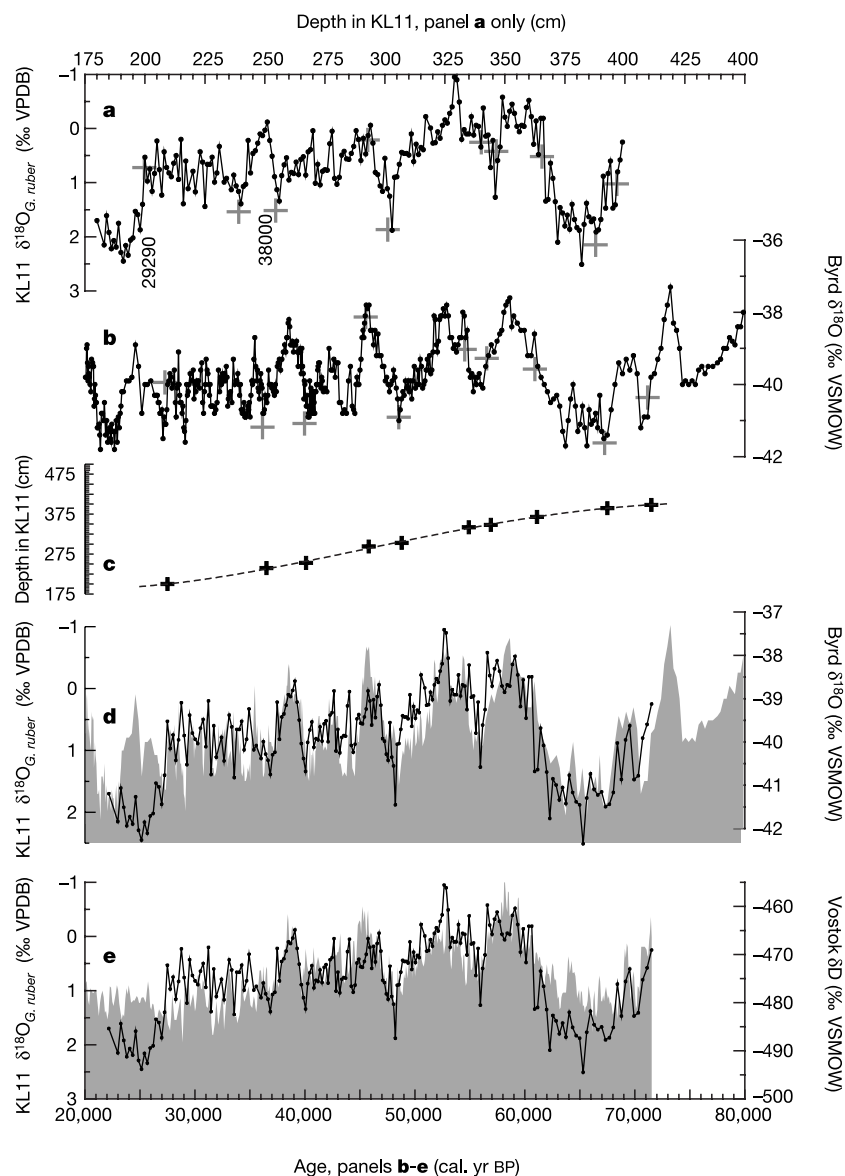


Figure 3 The high-resolution $\delta^{18}\text{O}$ record for the study interval between ~70 and ~20 kyr BP, based on analyses of the planktonic foraminifer *G. ruber* in central Red Sea core KL11^{19,20}. **a**, Original data versus depth in the core. Grey crosses indicate tie-points used to relate the chronology of KL11 to the West Antarctic Byrd ice-core $\delta^{18}\text{O}$ series³. Numbers shown are (AMS) ^{14}C datings, calibrated using a 400-yr reservoir age. **b**, The Byrd $\delta^{18}\text{O}$ record versus Greenland GISP2 ice-core equivalent age, based on synchronization of abrupt changes in the methane concentrations within the two ice cores³. Grey crosses indicate tie points as described for **a**. The youngest tie-point is

selected in a position nearest to the age suggested by the AMS ^{14}C dating in KL11. This also allows for the fact that the Byrd record is different from the East Antarctic Vostok δD series (**e**) in the interval ~25–17 kyr BP, possibly owing to a regional West Antarctic climate fluctuation². **c**, Depths of tie points in KL11 versus their equivalent age in the Byrd record. Dashed line represents a smooth fit through the tie-points for conversion of KL11 depths into age. **d**, KL11 $\delta^{18}\text{O}$ series after conversion to age as in **c**, superimposed upon the Byrd $\delta^{18}\text{O}$ series to illustrate the remarkable similarity in the signal structures. **e**, As **d** but using the Vostok δD record.

which—although not yet translated into absolute sea-level values—shows a good signal agreement with our sea-level reconstruction. The agreement is especially notable when the age model of KL11 is ‘tuned’ to that of independently dated MD952042 (Fig. 4b). The only realistic influences on the deep-sea $\delta^{18}\text{O}$ record of MD952042 are global ice volume (hence, sea level) and deep-sea temperature changes. A $\pm 1^\circ\text{C}$ uncertainty in deep-sea temperature variations implies a confidence limit of about $\pm 30\text{ m}$ on inferred sea levels, based on a 0.26‰ change in the water-to-calcite isotopic fractionation per 1°C , and a 0.008‰ change in Atlantic deep-sea $\delta^{18}\text{O}_{\text{water}}$ per metre sea-level lowering¹⁰. Offsets between the records (Fig. 4) that exceed $\pm 12\text{ m}$ around our reconstruction may be due to

analytical uncertainty, and/or deep-sea temperature changes affecting MD952042.

We have presented here an independent method for sea-level reconstruction that is applicable over timescales from glacial–interglacial to centennial. The reconstruction for the 70–25 kyr BP interval offered here is to our knowledge the first to be sufficiently resolved—in both the temporal sense and in terms of the confidence margin—to determine the realistic magnitudes and rates of sea-level shifts associated with millennial-scale climate fluctuations. We have corroborated an earlier report of signal similarity between the Antarctic temperature history and the sea-level record¹. Also, we have derived a magnitude of the sea-level shifts during the 70–25 kyr

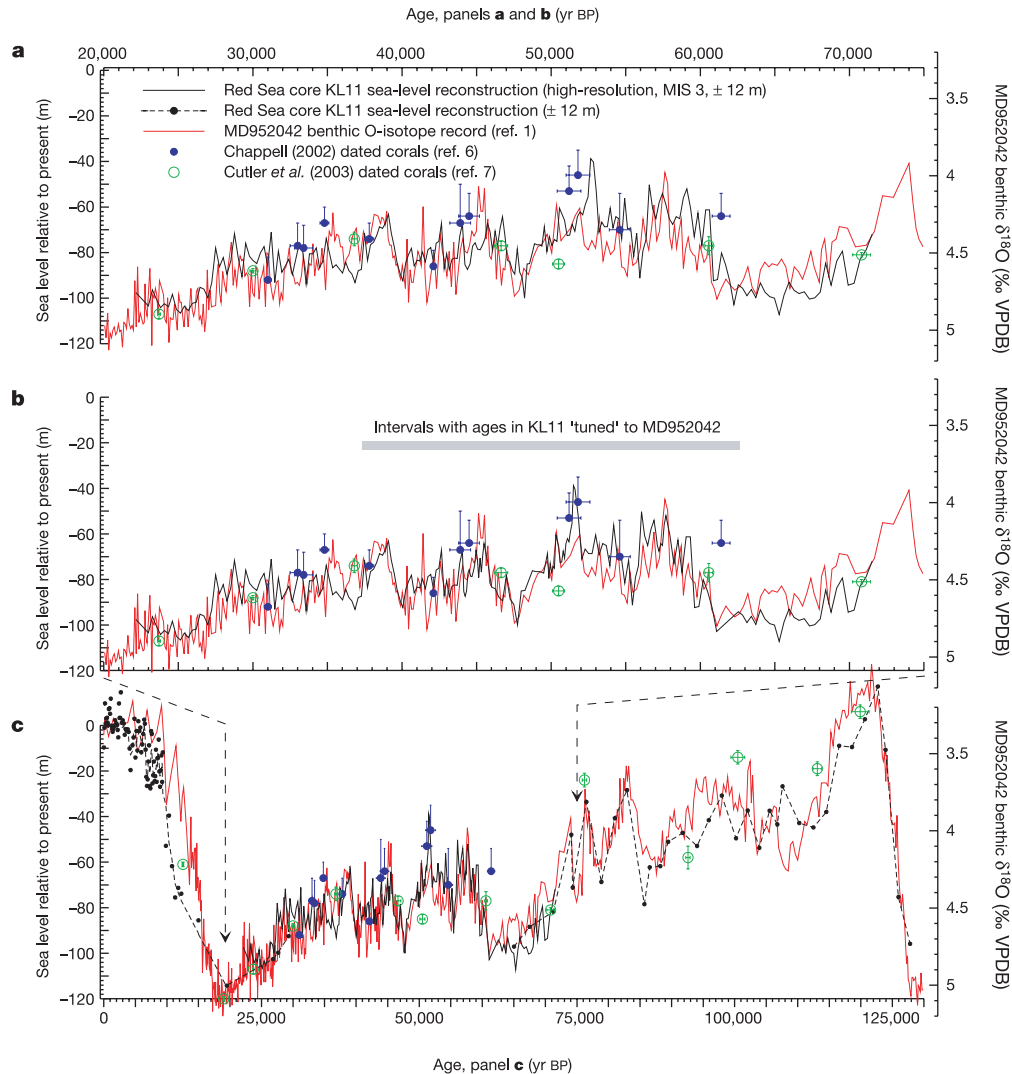


Figure 4 Comparison between the Red Sea and other sea-level estimates. **a**, The modelled high-resolution sea-level reconstruction for the period between 20 and 75 kyr BP from core KL11 (timescale as developed in Fig. 3). **b**, As **a** but with the KL11 timescale

fine tuned to the independently derived timescale of MD952042 (where possible on the basis of signal similarity). **c**, Red Sea sea level reconstruction between 0 and 130 kyr BP combining both high and low resolution records of core KL11.

BP interval of up to 35 ± 12 m, more than twice the volume of the Greenland and West Antarctic ice sheets. The maximum rate of change of $\sim 0.02 \text{ m yr}^{-1}$ is similar to the mean rate of change during the last deglaciation. □

Received 28 January; accepted 24 April 2003; doi:10.1038/nature01690.

- Shackleton, N. J., Hall, M. A. & Vincent, E. Phase relationships between millennial-scale events 64,000–24,000 years ago. *Paleoceanography* **15**, 565–569 (2000).
- Blunier, T. *et al.* Asynchrony of Antarctic and Greenland climate change during the last glacial period. *Nature* **394**, 739–743 (1998).
- Blunier, T. & Brook, E. Timing of millennial-scale climate change in Antarctica and Greenland during the last glacial period. *Science* **291**, 109–112 (2001).
- Dansgaard, W. *et al.* Evidence for general instability of past climate from a 250-kyr ice-core record. *Nature* **364**, 218–220 (1993).
- Groote, P. M., Stuiver, M., White, J. W. C., Johnsen, S. J. & Jouzel, J. Comparison of oxygen isotope records from the GISP2 and GRIP Greenland ice cores. *Nature* **366**, 552–554 (1993).
- Chappell, J. Sea-level changes forced ice breakouts in the last glacial cycle: New results from coral terraces. *Quat. Sci. Rev.* **21**, 1229–1240 (2002).
- Cutler, K. B. *et al.* Rapid sea-level fall and deep-ocean temperature change since the last interglacial period. *Earth Planet. Sci. Lett.* **206**, 253–271 (2003).
- Shackleton, N. J. Oxygen isotopes, ice volume and sea level. *Quat. Sci. Rev.* **6**, 183–190 (1987).
- Waelbroeck, C. *et al.* Sea-level and deep water temperature changes derived from benthic foraminifera isotopic records. *Quat. Sci. Rev.* **21**, 295–305 (2002).
- Adkins, J. F., McIntyre, K. & Schrag, D. P. The salinity, temperature, and delta O-18 of the glacial deep ocean. *Science* **298**, 1769–1773 (2002).

- Sofianos, S. S., Johns, W. & Murray, S. P. Heat and freshwater budgets in the Red Sea from direct observations at Bab el Mandab. *Deep Sea Res. II* **49**, 1323–1340 (2002).
- Rohling, E. J. *et al.* Magnitudes of sea-level lowstands of the past 500,000 years. *Nature* **394**, 162–165 (1998).
- Siddall, M., Smeed, D. A., Matthiesen, S. & Rohling, E. J. Modelling the seasonal cycle of the exchange flow in Bab el Mandab (Red Sea). *Deep Sea Res. I* **49**, 1551–1569 (2002).
- Griffiths, R. W. & Hopfinger, E. J. The structure of mesoscale turbulence and horizontal spreading at ocean fronts. *Deep Sea Res.* **31**, 245–269 (1984).
- Rohling, E. J. Environmental control on Mediterranean salinity and $\delta^{18}\text{O}$. *Paleoceanography* **14**, 706–715 (1999).
- Smeed, D. A. Hydraulic control of three-layer exchange flows: Application to the Bab-al-Mandab. *J. Phys. Oceanogr.* **30**, 2574–2588 (2000).
- Smeed, D. A. Seasonal variation of the flow in the strait of Bab al Mandab. *Oceanol. Acta* **20**, 773–781 (1997).
- Edwards, F. J. in *Key Environments: Red Sea* (eds Edwards, A. J. & Head, S. M.) 45–69 (Oxford, Pergamon, 1987).
- Hemleben, C. *et al.* Three hundred and eighty thousand year-long stable isotope and faunal records from the Red Sea. *Paleoceanography* **11**, 147–156 (1996).
- Schmelzer, L. *High-frequency Event Stratigraphy and Paleoclimatology of the Red Sea* Thesis, Eberhard-Karls-Univ. (1998).
- Fairbanks, R. G. The age and origin of the “Younger Dryas climate event” in Greenland ice cores. *Paleoceanography* **5**, 937–948 (1990).
- Bard, E., Hamelin, B., Fairbanks, R. G. & Zindler, A. Calibration of the ^{14}C timescale over the past 30,000 years using mass spectrometric U–Th ages from Barbados corals. *Nature* **345**, 405–409 (1990).
- Edwards, R. L. *et al.* A large drop in atmospheric ^{14}C and reduced melting in the Younger Dryas documented by ^{230}Th ages of corals. *Science* **260**, 962–968 (1993).

24. Bard, E. *et al.* Deglacial sea-level record from Tahiti corals and the timing of global meltwater discharge. *Nature* **382**, 241–244 (1996).
25. Fenton, M., Geiseler, S., Rohling, E. J. & Hemleben, C. A planktonic zones in the Red Sea. *Mar. Micropaleontol.* **40**, 277–294 (2000).
26. Almogi-Labin, A. *et al.* The influence of the NE winter monsoon on productivity changes in the Gulf of Aden, NW Arabian Sea, during the last 530 ka as recorded by foraminifera. *Mar. Micropaleontol.* **40**, 295–319 (2000).
27. Voelker, A. H. L. *et al.* Correlation of marine ^{14}C ages from the Nordic Seas with the GISP2 isotope record: Implications for radiocarbon calibration beyond 25 ky BP. *Radiocarbon* **40**, 517–553 (1998).
28. Pirazzoli, P. A. *et al.* A one million year sequence of marine terraces on Sumba Island, Indonesia. *Mar. Geol.* **109**, 221–236 (1993).
29. Gallup, C. D., Edwards, R. L. & Johnson, R. G. The timing of high sea levels over the past 200,000 years. *Science* **263**, 796–800 (1994).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank N. Shackleton for advice and permission to use the deep-sea $\delta^{18}\text{O}$ record of MD952042b, E. Bard for Tahiti sea-level estimates, and A. Hogg, J. F. McManus, P. M. Grootes and H. Erlenkeuser for help during the preparation of this Letter. M.S. has been funded by a NERC studentship.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.S. (ms14@soc.soton.ac.uk).

Archaean ultra-depleted komatiites formed by hydrous melting of cratonic mantle

A. H. Wilson^{*}, S. B. Shirey[†] & R. W. Carlson[†]

^{*} School of Geological and Computer Sciences, University of Natal, Durban 4041, South Africa

[†] Department of Terrestrial Magnetism, Carnegie Institution of Washington, 5241 Broad Branch Road, Washington DC 20015, USA

Komatiites are ultramafic volcanic rocks containing more than 18 per cent MgO (ref. 1) that erupted mainly in the Archaean era (more than 2.5 gigayears ago). Although such compositions occur in later periods of Earth history (for example, the Cretaceous komatiites of Gorgona Island²), the more recent examples tend to have lower MgO content than their Archaean equivalents. Komatiites are also characterized by their low incompatible-element content, which is most consistent with their generation by high degrees of partial melting (30–50 per cent³). Current models for komatiite genesis include the melting of rock at great depth in plumes of hot, diapirically rising mantle⁴ or the melting of relatively shallow mantle rocks at less extreme, but still high, temperatures caused by fluxing with water⁵. Here we report a suite of ultramafic lava flows from the Comondale greenstone belt, in the southern part of the Kaapvaal Craton, which represents a previously unrecognized type of komatiite with exceptionally high forsterite content of its igneous olivines, low $\text{TiO}_2/\text{Al}_2\text{O}_3$ ratio, high silica content, extreme depletion in rare-earth elements and low Re/Os ratio. We suggest a model for their formation in which a garnet-enriched residue left by earlier cratonic volcanism was melted by hydration from a subducting slab.

The well-preserved Comondale suite of lava flows, dated by whole-rock Sm–Nd at $3,334 \pm 18$ Myr age⁶, contains spinifex textures and aphyric chill margins representing liquid compositions. The 600-m thick succession, comprising more than 80 flow units, has been intersected by diamond drilling, allowing complete units to be studied with unambiguous recognition of

chill zones. Olivine and orthopyroxene are preferentially concentrated in the centres of the flows and both olivine and orthopyroxene spinifex occur near the top of the units, often in layered form. Olivine core compositions (higher than 96.5% forsterite content) have the highest magnesium content recorded for olivines in any naturally occurring terrestrial igneous rock. An abundance of primary liquidus orthopyroxene shows that these komatiites are different from all others previously described.

The Comondale liquids (Fig. 1) have high MgO (29–31%), high SiO_2 (49.5–50.5%) and extremely low iron contents (3.5–5% FeO). They are characterized by exceptionally low levels of incompatible elements (TiO_2 0.1%; Zr 2.6 p.p.m.; Y 5.2 p.p.m.; total rare-earth elements (REE) 2.9 p.p.m.) compared with other komatiites (TiO_2 0.3–1.61%; Zr 14–27 p.p.m.; Y 5–11 p.p.m.; total REE 9–19 p.p.m.). They also have the highest values for $\text{Al}_2\text{O}_3/\text{TiO}_2$ (65–95) recorded for any komatiite and exceptionally low Zr/Y (0.5 compared to 3.3 for Barberton-type komatiites). The latter is significant in that it indicates melting of a garnet-enriched mantle source. These lavas exhibit much greater light-REE depletion than the previously known most-extreme compositions, those of the Cretaceous komatiites of Gorgona Island (Fig. 2). Chondrite normalized values for Gd/Yb and La/Yb are 0.29 and 0.02 respectively, compared with 1–1.3 and 0.3–1.6 for other komatiites, including Barberton. The Comondale lavas are derived from a highly depleted and refractory mantle source of a type not previously recognized.

Comondale komatiites have extremely low Re (0.02–0.09 p.p.b.) and high Os (1–2 p.p.b.) contents that lead to low $^{187}\text{Re}/^{188}\text{Os}$ with a limited spread in Os isotopic composition (Fig. 3; Table 1). Although the samples show good linearity in the Re–Os isochron diagram, the limited range in the Re/Os produces a large uncertainty (440 Myr) about an age of 3,393 Myr, but this age agrees with the precise Sm–Nd isochron ($3,334 \pm 18$ Myr age⁶). The low Re/Os of the samples tightly constrains the $^{187}\text{Os}/^{188}\text{Os}$ of their mantle source to 0.1047 ± 0.0017 ($\gamma_{\text{Os}} = 0.6 \pm 1.6$). This is equivalent to a source with a time-averaged chondritic Re/Os at 3,334 Myr ago. These Os isotopic compositions provide the best estimate until now of the Earth's oldest mantle composition obtained directly from a melt. The data confirm that the Earth's mantle had evolved from 4.5 to 3.3 Gyr ago with a chondritic time-averaged Re/Os and strongly supports a late chondritic veneer for highly siderophile elements^{7–10}. A source of this composition gives no evidence of incorporation of radiogenic Os from the crust, a conclusion supported by the high initial Nd isotopic composition ($\epsilon_{\text{Nd}} = +2.0 \pm 0.7$)⁶ and its extreme light-REE depletion. These

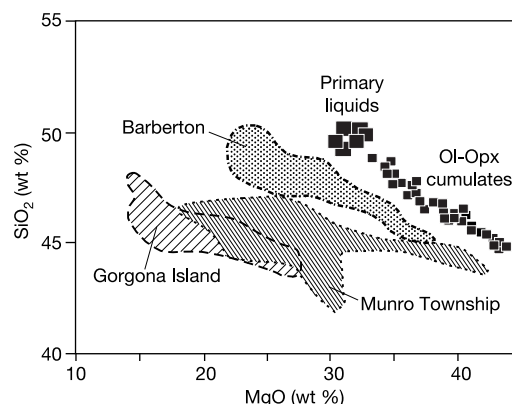


Figure 1 Variation of SiO_2 and MgO for Comondale liquid and olivine–orthopyroxene cumulates and for other komatiites. These are: Barberton komatiites²⁰; komatiites from Munro Township (Abitibi belt)^{21,22}; komatiites and picrites from Gorgona Island². Ol, olivine; Opx, orthopyroxene.

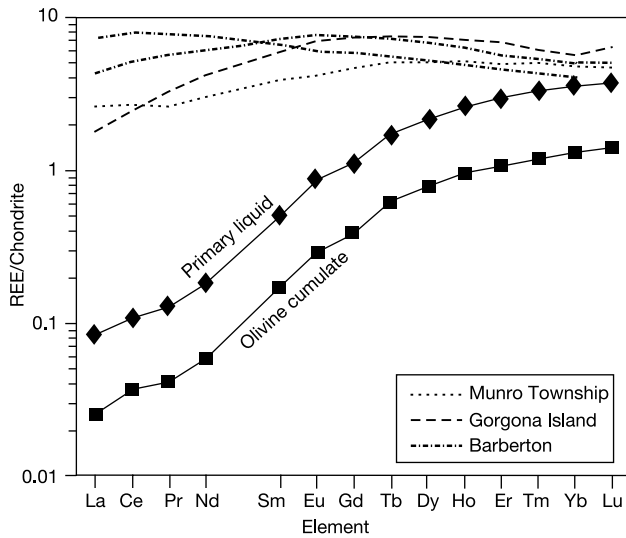


Figure 2 Chondrite normalized rare-earth patterns for liquid and olivine cumulate compositions for Commondale komatiites. Other komatiite liquid compositions are for Gorgona Island², Munro Township^{21,22}, Barberton³ and this study (curve with lower La content).

features place strong constraints on the timing of the depletion event experienced by the source of the magmas. Assuming Sm/Nd of the parental magma (0.884) would reflect that of the mantle source at such large degrees of melting, the ϵ_{Nd} of this source would increase by over 4 ϵ units every 100 Myr. It is not likely that the magma source of the Commondale magmas started with an exceptionally low $^{143}\text{Nd}/^{144}\text{Nd}$, outside the range typical of early Archaean mantle-derived magmas with little crustal contamination ($\epsilon_{\text{Nd}} = 0$ to +4). Thus the extreme depletion of the Commondale source could not have occurred more than 50–100 Myr before the eruption of the Commondale magmas and must have occurred during the time interval of the growth of the oldest Kaapvaal lithosphere.

Most komatiites in the geological record crystallized augite, in some cases together with pigeonite. The dominance of orthopyroxene in the Commondale komatiites (and some boninites, a modern rock type that is unique to subduction zones¹¹) reflects the highly depleted (in high-Ca pyroxene component) and silica-rich character of these magmas (Fig. 4a). The rocks do not contain pigeonite because the liquid line of descent did not extend sufficiently far before the liquids were quenched. This is consistent with the observed low-pressure sequence of crystallization of olivine, followed by olivine and orthopyroxene. The link between such unique composition ultramafic lava flows and the melting regime in the mantle is complex but melt compositions and melting experiments carried out on likely mantle materials and modelling calculations on ultramafic melts produced from fertile and depleted mantle peridotite¹² provide some constraint. Projections of the

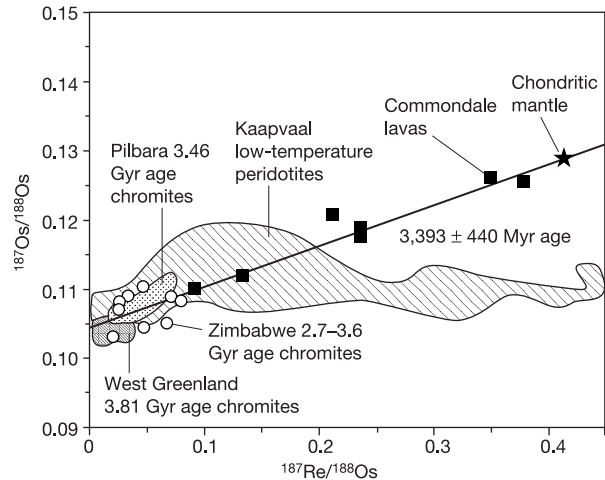


Figure 3 $^{187}\text{Re}/^{188}\text{Os}$ versus $^{187}\text{Os}/^{188}\text{Os}$ for the Commondale komatiites. The figure emphasizes the link between these lavas and typically depleted lithosphere xenoliths from the sub-continental mantle^{23,24} in the Kaapvaal Craton erupted in kimberlite. The age ($3,393 \pm 440$ Myr) is within error of the accepted $3,334 \pm 18$ Myr age⁶ on the precise Sm–Nd whole-rock isochron for Commondale and implies a mantle source with a chondritic Os isotopic composition at the eruption age. Commondale lavas have an initial Re/Os isotopic composition comparable to those of early Archaean chromites from Greenland²⁵, Zimbabwe²⁶ and the Pilbara²⁵ cratons.

Commondale liquid compositions and olivine accumulates on both the Mg–Tschermak’s–olivine–enstatite¹² (Fig. 4b) and on the anorthite–olivine–quartz projections (not shown)¹² show that the Commondale liquids lie in the field of melting of harzburgite at 3.5–4 GPa. Such melts generated in harzburgitic mantle are entirely consistent with the observed low-pressure crystallization of olivine and orthopyroxene.

This compares with projections for aphyric (and therefore most likely to represent liquid compositions) komatiites from Barberton, which indicate pressures of 8–10 GPa (ref. 13). The depth of Commondale komatiite generation is indicated to be relatively shallow compared to that of Barberton. The melting conditions that generated komatiites from Gorgona Island are also not applicable to the Commondale komatiites, because, even though they involved similar pressures of melt generation^{2,12,13}, the liquidus phase assemblages, compositions and trace element character are significantly different (Figs 1, 2 and 4), showing that either the source material or melting model as used for Gorgona Island is not appropriate for the Commondale rocks.

Although a plume model^{4,14} may explain the Commondale komatiites, such a plume would have had very unusual compositional characteristics. For example, a paradox exists for the Commondale komatiites in that the signature of garnet melting in the source (high Y/Zr, low Gd/Ybn, high Al_2O_3), as appropriate for a mantle residue after extraction of Barberton-type komatiite magma, must be reconciled against the relative low pressure of

Table 1 Re–Os isotopic data for Commondale lavas

Sample	Re (p.p.b.)	Os (p.p.b.)	$^{187}\text{Re}/^{188}\text{Os}$	$^{187}\text{Os}/^{188}\text{Os}$	$^{187}\text{Os}/^{188}\text{Os}$ (initial ratio, 3.33 Gyr ago)	γ_{Os} (3.33 Gyr ago)
BH2 39.46	0.0203(6)	1.041(3)	0.0936(28)	0.11171(47)	0.1064(3)	2.21(6)
BH2 43.53	0.0280(8)	1.480(4)	0.0883(26)	0.10961(11)	0.1046(3)	0.48(6)
BH2 57.55	0.0668(20)	1.386(4)	0.2334(70)	0.11838(21)	0.1051(4)	0.95(7)
BH2 78.21	0.0378(11)	2.220(7)	0.0826(25)	0.10880(11)	0.1041(3)	0.01(6)
BH2 91.08	0.0541(16)	1.246(4)	0.2107(63)	0.12092(96)	0.1089(4)	4.63(7)
COM 1–4	0.0609(18)	1.254(4)	0.2337(70)	0.11784(17)	0.1045(4)	0.41(7)
COM 1–20	0.0438(13)	1.592(5)	0.1323(40)	0.11179(21)	0.1042(3)	0.16(6)
COM 1–25	0.0780(23)	1.102(5)	0.3770(113)	0.12568(7)	0.1042(4)	0.09(7)
COM 1–7	0.0959(29)	1.329(4)	0.3476(104)	0.12611(25)	0.1063(4)	2.11(7)

Uncertainties at 2σ applicable to the last digits are given in parentheses.

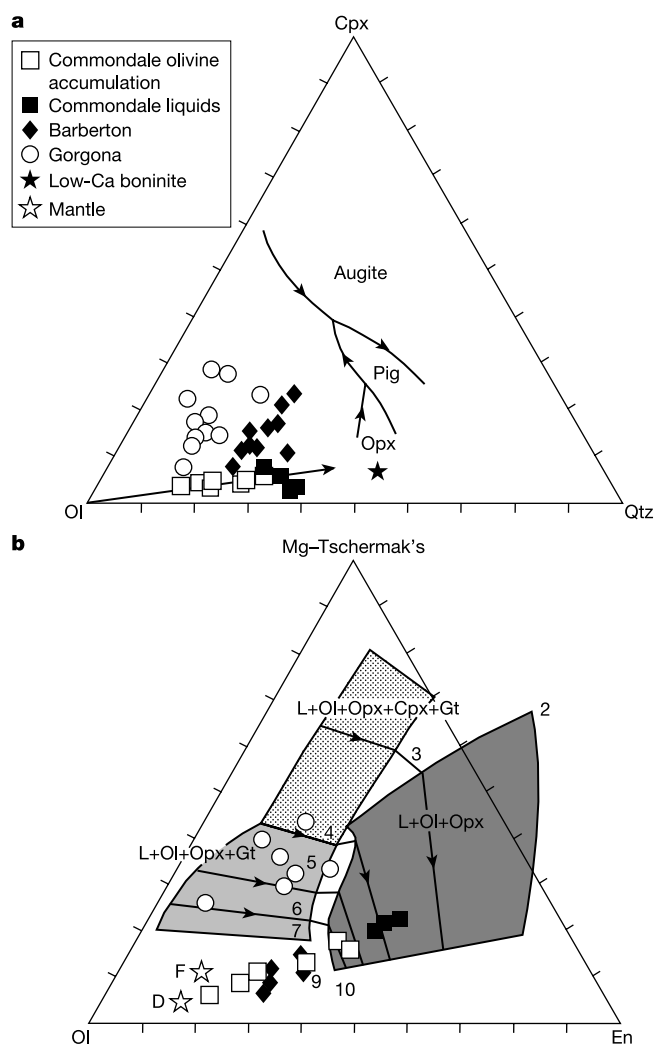


Figure 4 Projections of Comondale liquid compositions and olivine accumulates into two systems. **a**, The olivine-clinopyroxene-quartz system²⁷; and **b**, Mg-Tschermak's-olivine-enstatite system¹². Augite is absent in the Comondale rocks because the liquid line of descent intersects the olivine-orthopyroxene boundary far down from the augite stability field. Shown also for comparison are komatiite compositions from Gorgona Island², aphyric komatiites from the Barberton greenstone belt²⁰, and a low-Ca boninite²⁸ Cpx, clinopyroxene; En, enstatite; Opx, orthopyroxene; Ol, olivine; L, liquid; Pig, pigeonite; Gt, garnet; Qtz, quartz. F, D, fertile and depleted mantle. The numbers from 2 to 10 indicate cotectic equilibrium pressures in GPa, and the arrows point up-temperature in the sense of advanced partial melting.

melting as indicated by phase diagrams. Furthermore, no known plume has generated primary lavas of such unique high-silica, high-Mg character (Fig. 1).

The role of water in melting to form komatiites versus anhydrous mantle upwelling has received considerable attention⁵. The question has not yet been resolved by experimental petrology, but water in a mantle source during melting at relatively shallow levels is known to decrease the solidus, allowing refractory mantle to melt at lower temperatures and to increase the silica content of the resultant melt^{15,16}. Since the Comondale source was highly depleted in incompatible elements (including water), if water was necessary for the generation of these magmas, it probably was introduced into the source by a second-stage process, such as subduction. Subduction processes are well established for the mid- to late-Archaean (2.9–2.5 Gyr ago)¹⁷ in Canada, Australia and Zimbabwe^{17,18}, but similar convincing evidence is lacking for the early Archaean because

exposures of rocks of this age are considerably more sparse and discontinuous and are often pervasively deformed.

A possible scenario for the Comondale komatiites is that the FeO-poor Barberton-type residue, enriched in garnet from the previous melting event, may have become stabilized at a level considerably shallower than the depth of previous melt extraction. A hot, buoyant and FeO-depleted residue would have been expected to rise dynamically into the shallow mantle regardless of the setting of initial melting, eventually to be incorporated as cratonic mantle. A mantle portion rising along an adiabat would cause solution of residual garnet and clinopyroxene into orthopyroxene and the shallow residue would be olivine ± orthopyroxene ± garnet. Hydration of this material at shallower levels could then have occurred by fluid released from the breakdown of hydrous minerals in a nearby subducting slab at approximately 100 km depth¹⁹. The Comondale suite has certain compositional characteristics in common with low-Ca boninites that are typically generated in subduction zones (high SiO₂, high MgO, low TiO₂ and Zr, heavy-REE depletion relative to most komatiites), but clearly originated from a much more depleted mantle source.

We suggest a hybrid model for Comondale komatiites that involved shallow melting and an introduction of water into the melting regime. This caused the previously depleted but newly hydrated residue to undergo a second stage of melting to form ultra-depleted magma of the type observed in the Comondale magmas. The Comondale mantle source has a composition close to that expected for pre-metasomatized Kaapvaal lithospheric mantle. The suggested petrogenetic model for the Comondale komatiites provides an important link between Archaean komatiite genesis and cratonic lithosphere stabilization. □

Received 28 November 2002; accepted 22 April 2003; doi:10.1038/nature01701.

- Kerr, A. C. & Arndt, N. T. A note on the IUGS reclassification of the high-Mg and picritic volcanic rocks. *J. Petrol.* **42**, 2169–2171 (2001).
- Révilion, S., Arndt, N. T., Chauvel, C. & Hallot, E. Geochemical study of ultramafic volcanic and plutonic rocks from Gorgona Island, Colombia: plumbing system of an oceanic plateau. *J. Petrol.* **41**, 1127–1153 (2000).
- Sun, S.-S. & Nesbitt, R. W. Petrogenesis of Archaean ultrabasic and basic volcanics: evidence from the rare earth elements. *Contrib. Mineral. Petrol.* **65**, 301–325 (1978).
- Campbell, I. H., Griffiths, R. W. & Hill, R. I. Melting in an Archaean mantle plume: heads it's basalts, tails it's komatiites. *Nature* **339**, 697–699 (1989).
- Parman, S. W., Dann, J. C., Grove, T. L. & de Wit, M. J. Emplacement conditions of komatiite magmas from the 3.49 Ga Komati Formation, Barberton greenstone belt, South Africa. *Earth Planet. Sci. Lett.* **150**, 303–323 (1997).
- Wilson, A. H. & Carlson, R. W. A Sm-Nd and Pb-Pb isotopic study of Archaean greenstone belts in the southern Kaapvaal Craton, South Africa. *Earth Planet. Sci. Lett.* **96**, 89–105 (1989).
- Meisel, T., Walker, R. J. & Morgan, J. W. The osmium isotopic composition of the Earth's primitive upper mantle. *Nature* **383**, 517–520 (1998).
- Morgan, J. W. Osmium isotopic constraints on Earth's late accretionary history. *Nature* **317**, 703–705 (1985).
- Morgan, J. W. Ultramafic xenoliths: Clues to the Earth's late accretionary history. *J. Geophys. Res.* **91**, 12375–12387 (1986).
- Shirey, S. B. & Walker, R. J. The Re-Os isotopic system in cosmochemistry and igneous geochemistry. *Annu. Rev. Earth Planet. Sci.* **26**, 425–500 (1998).
- Crawford, A. J., Falloon, T. J. & Green, D. H. in *Boninites and Related Rocks* (ed. Crawford, A. J.) 2–44 (Unwin and Hyman, London, 1989).
- Herzberg, C. & O'Hara, M. J. Plume-associated ultramafic magmas of Pherozoic age. *J. Petrol.* **43**, 1857–1883 (2002).
- Walter, M. J. Melting of garnet peridotite and the origin of komatiite and depleted lithosphere. *J. Petrol.* **39**, 29–60 (1998).
- Campbell, I. H. & Griffiths, R. W. Implications of mantle plume structure for the evolution of flood basalts. *Earth Planet. Sci. Lett.* **99**, 79–93 (1990).
- Kawamoto, T. & Holloway, J. R. Melting temperature and partial melt chemistry of H₂O saturated mantle peridotite to 11 gigapascals. *Science* **276**, 240–243 (1997).
- Kushiro, I. Effect of water on the composition of magmas formed at high pressures. *J. Petrol.* **13**, 311–334 (1972).
- de Wit, M. J. On Archaean granites, greenstones, cratons and tectonics: does the evidence demand a verdict? *Precamb. Res.* **91**, 181–226 (1998).
- Zegers, T. E., White, S. H., de Keijzer, M. & Dirks, P. Extensional structures during the deposition of the 3460 Ma Warrawoona Group in the eastern Pilbara Craton, Western Australia. *Precamb. Res.* **80**, 89–105 (1996).
- Schmidt, M. W. & Poli, S. Experimentally based water budgets for dehydrating slabs and consequences for arc magma generation. *Earth Planet. Sci. Lett.* **163**, 361–379 (1998).
- Smith, H. S. & Erlank, A. J. in *Komatiites* (eds Arndt, N. T. & Nisbet, E. G.) 347–398 (George Allen and Unwin, London, 1982).
- Fan, J. & Kerrich, R. Geochemical characteristics of Al-depleted and Al-undepleted komatiites and

- HREE-enriched low-Ti tholeiites, western Abitibi greenstone belt: A heterogeneous mantle plume-convergent margin environment. *Geochim. Cosmochim. Acta* **61**, 4723–4744 (1997).
22. Xie, Q. & Kerrich, R. Silicate-perovskite and marjorite signature komatiites from the Archean Abitibi Greenstone Belt: implications for early mantle differentiation and stratification. *J. Geophys. Res.* **99**, 15799–15812 (1994).
 23. Carlson, R. W. et al. in *The J.B. Dawson Volume—Proc. 7th Int. Kimberlite Conf* (eds Gurney, J. J., Gurney, J. L., Pascoe, M. D. & Richardson, S. H.) 99–108 (Red Roof Design, Cape Town, 1999).
 24. Irvine, G., Pearson, D. G. & Carlson, R. W. Lithospheric mantle evolution of the Kaapvaal Craton: a Re-Os isotopic study of peridotite xenoliths from Lesotho kimberlites. *Geophys. Res. Lett.* **28**, 2505–2508 (2001).
 25. Bennett, V. C., Nutman, A. P. & Esat, T. M. Constraints on mantle evolution from $^{187}\text{Os}/^{188}\text{Os}$ isotopic compositions of Archean ultramafic rocks from southern West Greenland (3.8 Ga) and Western Australia (3.46 Ga). *Geochim. Cosmochim. Acta* **66**, 2615–2630 (2002).
 26. Nagler, T. F., Kramers, J. D., Kamber, B. S., Frei, R. & Prendergast, M. D. Growth of subcontinental lithospheric mantle beneath Zimbabwe started at or before 3.8 Ga: Re-Os study on chromites. *Geology* **25**, 983–986 (1997).
 27. Kinzler, R. J. & Grove, T. L. Crystallization and differentiation of Archean komatiite lavas from northeast Ontario: phase equilibrium and kinetic studies. *Am. Mineral.* **70**, 40–51 (1985).
 28. Taylor, R. N. et al. Mineralogy, chemistry, and genesis of the boninite series volcanics, Chichijima, Bonin Islands, Japan. *J. Petrol.* **35**, 577–617 (1994).

Acknowledgements Support for this study comes from a NSF EAR Continental Dynamics grant, the National Research Foundation (South Africa) and the University of Natal Research Fund. C. Herzberg and M. Walter provided comments that improved this paper. We also thank N. Arndt, M. Cheadle, J. Boyd, S. Parman, M. de Wit and B. Mysen for discussions on these komatiites, but the views expressed in this paper are our own.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.H.W. (wilsona@nu.ac.za).

Automatic gain control in the echolocation system of dolphins

Whitlow W. L. Au & Kelly J. Benoit-Bird

Hawaii Institute of Marine Biology, University of Hawaii, PO Box 1106, Kailua, Hawaii 96734, USA

In bats¹ and technological sonars², the gain of the receiver is progressively increased with time after the transmission of a signal to compensate for acoustic propagation loss. The current understanding of dolphin echolocation indicates that automatic gain control is not a part of their sonar system³. In order to test this understanding, we have performed field measurements of free-ranging echolocating dolphins. Here we show that dolphins do possess an automatic gain control mechanism, but that it is implemented in the transmission phase rather than the receiving phase of a sonar cycle. We find that the amplitude of the dolphins' echolocation signals are highly range dependent; this amplitude increases with increasing target range, R , in a $20 \log(R)$ fashion to compensate for propagation loss. If the echolocation target is a fish school with many sound scatterers, the echoes from the school will remain nearly constant⁴ with range as the dolphin closes in on it. This characteristic has the same effect as time-varying gain in bats and technological sonar when considered from a sonar system perspective.

A well-established method to increase the dynamic range of sonar is to vary the gain of the receiver as a function of time from the transmission of a sonar pulse—a technique referred to as time-varying gain, or TVG². The gain of the receiver is programmed to increase with time to compensate for a one-way spherical spreading loss of $20 \log(R)$, or for a two-way loss of $40 \log(R)$. The gain is usually increased in a $20 \log(c\Delta t)$ or a $40 \log(c\Delta t)$ fashion (where c is the speed of sound and Δt is the time elapsed from the time of transmission, so that $c\Delta t$ is equivalent to R), until a specific maximum gain level is reached.

The auditory system of the bat *Eptesicus fuscus* performs a process that is analogous to TVG. The bat's hearing sensitivity decreases at the moment of a sonar signal emission and then begins to increase at a rate of 12 dB per doubling of the elapsed time, which is equivalent to a $40 \log(c\Delta t)$ response. This process continues until an elapsed time of approximately 6.4 ms, corresponding to a target range of 1.4 m, after which the bat regains its normal hearing sensitivity¹. Therefore as the bat closes within 1.4 m of its prey, the amplitude of the echoes remains constant if the transmission level remains constant. The reduced sensitivity is the result of the bat's middle-ear muscles contracting in synchrony with a vocalization to protect the auditory system from the emitted sound. The muscle contraction actually begins before the onset of vocalization^{5,6}. Such an automatic gain control phenomenon has not been observed with dolphins⁷. However, recent echolocation signal measurements of free-ranging dolphins suggest that their echolocation system also possesses a gain control capability.

Accurate measurements of echolocation signals used by free-ranging dolphins in the wild can be difficult to obtain, because the echolocation beam pattern is relatively narrow⁷. If echolocation signals are not measured close to the axis of the animal's beam, the signals will be distorted⁷. It is also extremely difficult to determine the distance of a moving dolphin from the recording hydrophone in order to determine the source level (sound pressure level 1 m from the dolphin) of the signals. We have successfully overcome these problems by using a short-base-line array of four hydrophones arranged as a symmetrical star, shown schematically in Fig. 1^{3,8,9}. The array structure resembles the letter "Y", with each 45.7-cm-long arm separated by an angle of 120° . A spherical hydrophone is connected to the end of each arm, with a fourth hydrophone connected at the geometric centre of the "Y." Each hydrophone output is connected to a simultaneous sampling four channel analog-to-digital converter controlled by a transportable PC aboard a boat. Calibration measurements obtained with the array and a simulated dolphin signal indicated only a 12% error in the range estimation at 25 m, which translated to a propagation loss error of only 1.1 dB (ref. 3).

The symmetrical star array has been used to measure the echolocation signals of the Hawaiian spinner dolphin (*Stenella longirostris*), pan-tropical spotted dolphins (*Stenella attenuata*), Atlantic spotted dolphins (*Stenella frontalis*), white-beaked dolphins (*Lagenorhynchus albirostris*), and killer whales (*Orcinus orca*). With the exception of the killer whale, echolocating dolphins in the wild typically emit broadband, short-duration click signals ($<80 \mu\text{s}$) with a bimodal spectrum; a low-frequency peak occurs between 30 and 50 kHz, and a high-frequency peak between 80 and 120 kHz. Approximately 70–80% of the signals measured in the field have been bimodal. Killer whales emit echolocation signals that are

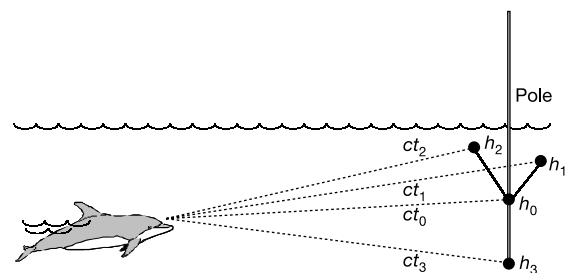


Figure 1 A schematic diagram of the symmetrical star array, with a hydrophone attached to the end of each arm and one at the centre of the array. The range of the dolphin to each hydrophone is denoted as ct_i , where c is the speed of sound in water and t_i is the signal propagation time from the dolphin to the i th hydrophone. Range was determined by measuring the difference in the time of arrival of a signal at each hydrophone.

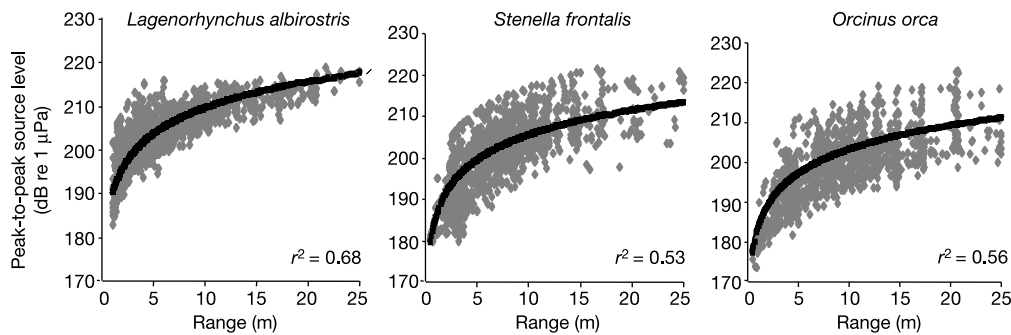


Figure 2 Source levels of three odontocete species measured in the wild as a function of the range of the animals from the symmetrical star array. The solid curve in each graph

represents the function $K + [20 \log(R)]$, where K is a constant that results in the best fit of the function to the measured data.

twice as long in duration and with frequency content approximately half that of the other dolphins studied.

During field measurements, dolphins often milled about 30–50 m from our boat and made ‘runs’ towards the array, continuously echolocating as they approached. The range of an echolocating dolphin was determined by measuring the time of arrival differences between the signal at the centre and the three other hydrophones (Fig. 1). Only signals in which the centre hydrophone detected the largest amplitude, or in which the central hydrophone detected signals that were within 3 dB of the largest amplitude, were analysed. This ensured that the dolphin’s echolocation beam was directed towards the array.

In Fig. 2 we show the peak-to-peak source level of the echolocation signals of *L. albirostris*, *S. frontalis* and *O. orca* as a function of range. The data points are fitted with a function $K + [20 \log(R)]$, where K is a constant that provides the best fit between the function and the measured data. The $20 \log(R)$ term represents the one-way transmission loss of the echolocation signal for a given R . The echolocation signals of all the dolphin species measured with our array displayed a variation with range that can be fitted well with a $20 \log(R)$ curve. The results indicate that as dolphins close in on a target, the source level of the echolocation signal decreases continuously by 6 dB for every halving of the range. This reduction in source level with decreasing range to the array is a form of dynamic time-varying gain control for the dolphin’s echolocation system. Instead of the receiving gain being manipulated, the output level of the system is manipulated. In conventional sonar systems, the amplitude of the emitted signal is held constant while TVG is applied to the receiver. For the bat, the emitted signal also tends to be relatively constant close to the terminal phase of prey capture, and the hearing sensitivity is manipulated via a middle-ear muscle contraction¹. The middle ear of dolphins is extremely complex, and its role in hearing and how it functions is unclear¹⁰. The ossicular chain is stiffened and tightly bound together with sheaths and annular ligaments, and the ossicles are denser and more massive than for terrestrial mammals of similar size¹⁰. These adaptations work against any effective middle-ear reflex, so that another means of gain control would be needed.

In order to maintain a constant echo level for a single prey item as a dolphin closes in, the gain of the sonar system should vary as $40 \log(R)$ rather than $20 \log(R)$. However, many dolphin species forage on fish schools¹¹. Sonar echoes from a fish school would be similar to volume reverberation, consisting of the sum of many individual echoes reflecting from individual fish. The amplitude of echoes from volume reverberation increases with decreasing range as a function of $20 \log(R)$ (ref. 1). Therefore, as a dolphin closes in on a fish school, the amplitude of the echoes will remain constant as the animal dynamically and progressively reduces the amplitude of the outgoing signal at a rate of $20 \log(R)$. When a dolphin closes in

on a single prey, the echo level will increase as the distance between dolphin and prey decreases, but not as rapidly as in a situation where there is no gain control.

The dynamic control of the echolocation source level is probably not the result of a cognitive process, but rather a natural consequence of how echolocation clicks are produced. Dolphins typically emit echolocation clicks at a rate that allow the echoes to return to the animal before the next click is emitted⁷. Consequently, the repetition rate increases as an animal closes on a target^{12,13}. The clicks are produced within the nasal system of the dolphins by manipulating the air flow through the phonic lips, previously referred to as the dorsal bursae/monkey lips complex^{14,15}. Dolphins initially pressurize their nasal system¹⁶ and then emit a click train, with the clicks occurring at relatively low repetition rate and the animal continually adjusting the rate as targets are located¹⁷. If the dolphin chooses to keep the amount of acoustic energy emitted relatively constant or within certain limits for each pressurization cycle, then the amplitude of the signal can be high when the repetition rate is low but must continually decrease as the repetition rate increases. The data in Fig. 2 are consistent with the notion that there is a coupling between repetition rate and source level.

An advantage of having a coupling between source level and target range, and subsequently click repetition rate, is that the relative size of a target can be inferred from the reception of a few echoes. The source level of echolocation signals is coupled to target range, so that for a given range, the source level is predictable within some small variation, making relatively simple the comparison of the level of a specific echo with levels of echoes from previous experience. The range of a target does not need to be accounted for, but is automatically factored in by the coupling of source level and repetition rate of the signals. At long ranges, scanning of the beam by the dolphin, an effective method at short ranges, will not be useful in determining the size of a target because the area a sonar beam covers can be much larger than the target. However, a dolphin can still determine the relative size of a target by the levels of the echoes. From a conventional sonar design perspective, the notion of time-varying gain is always associated with the receiver. However, if the receiver gain cannot be controlled, then an alternative is to control the level of the transmitted signal, as in the case of a dolphin. Working within a dolphin’s physiological constraints to maximize the efficiency of echolocation, the dolphin sonar system has arrived at a gain control system that is different from those of bats and technological sonars, but which is just as effective. □

Received 27 November 2002; accepted 8 April 2003; doi:10.1038/nature01727.

1. Simmons, J. A., Moffat, A. J. M. & Masters, W. M. Sonar gain control and echo detection thresholds in the echolocating bat, *Eptesicus fuscus*. *J. Acoust. Soc. Am.* **91**, 1150–1163 (1992).
2. Medwin, H. & Clay, C. S. *Fundamentals of Acoustical Oceanography* (Academic, San Diego, 1998).
3. Au, W. W. L. & Herzing, D. L. Echolocation signals of wild Atlantic spotted dolphin (*Stenella frontalis*). *J. Acoust. Soc. Am.* **113**, 598–604 (2003).

4. Urlick, R. J. *Principles of Underwater Sound* (McGraw Hill, New York, 1983).
5. Henson, O. W. Jr The activity and function of the middle ear muscles in echolocating bats. *J. Physiol. (Lond.)* **180**, 871–887 (1965).
6. Suga, N. & Jen, P. H. S. Peripheral control of acoustic signals in the auditory system of echolocating bats. *J. Exp. Biol.* **62**, 277–331 (1975).
7. Au, W. W. L. *The Sonar of Dolphins* (Springer, New York, 1993).
8. Au, W. W. L., Ford, J. K. B. & Allman, K. A. Echolocation signals of foraging killer whales (*Orcinus orca*). *J. Acoust. Soc. Am.* **111**, 2343–2344 (2002).
9. Rasmussen, M. H., Miller, L. A. & Au, W. W. L. Source levels of clicks from free-ranging white beaked dolphins (*Lagenorhynchus albirostris* Gray 1846) recorded in Icelandic waters. *J. Acoust. Soc. Am.* **111**, 1122–1125 (2002).
10. Ketten, D. R. in *Hearing by Whales and Dolphins* (eds Au, W. W. L., Popper, A. N. & Fay, R. R.) 43–108 (Springer, New York, 2000).
11. Heithaus, M. R. & Dill, L. M. in *Encyclopedia of Marine Mammals* (eds Perrin, W. F., Wursig, B. & Theewissen, J. G. M.) 411–422 (Academic, San Diego, 2002).
12. Evans, W. W. & Powell, B. A. in *Animal Sonar Systems: Biology and Bionics* (ed. Busnel, R. G.) 363–382 (Laboratoire de Physiologie Acoustique, Jouy-en-Josas, 1967).
13. Morozov, B. P., Akapiam, A. E., Burdin, V., Zaitseva, K. A. & Solovkh, Y. A. Tracking frequency of the location signals of dolphins as a function of distance to the target. *Biofizika* **17**, 139–145 (1972).
14. Cranford, T. in *Animal Sonar: Processes and Performance* (eds Nachtigall, P. E. & Moore, P. W. B.) 67–77 (Plenum, New York, 1988).
15. Cranford, T. in *Hearing by Whales and Dolphins* (eds Au, W. W. L., Popper, A. N. & Fay, R. R.) 109–155 (Springer, New York, 2000).
16. Ridgway, S. H. et al. Hearing and whistling in the deep sea: Depth influences whistle spectra but does not attenuate hearing by white whales (*Delphinapterus leucas*) (Odontoceti cetacea). *J. Exp. Biol.* **204**, 3829–3841 (2001).
17. Penner, R. H. in *Animal Sonar: Processes and Performance* (eds Nachtigall, P. E. & Moore, P. W. B.) 707–713 (Plenum, New York, 1988).

Acknowledgements We thank the following individuals for their assistance in collecting data in the field: D. Herzog (with *S. frontalis*), L. Miller and M. Rasmussen (with *L. albirostris*), and J. Ford and K. Allman (with *O. orca*). This work was supported, in part, by the Office of Naval Research.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to W.A. (wau@hawaii.edu).

Involvement of *Notch* and *Delta* genes in spider segmentation

Angelika Stollewerk*, Michael Schoppmeier* & Wim G. M. Damen

Institute for Genetics, Evolutionary Genetics, University of Cologne, Weyertal 121, D-50931 Köln, Germany

* These authors contributed equally to this work

It is currently debated whether segmentation in different animal phyla has a common origin and shares a common genetic mechanism^{1,2}. The apparent use of different genetic networks in arthropods and vertebrates has become a strong argument against a common origin of segmentation. Our knowledge of arthropod segmentation is based mainly on the insect *Drosophila*, in which a hierarchical cascade of transcription factors controls segmentation^{3,4}. The function of some of these genes seems to be conserved among arthropods, including spiders^{5,6}, but not vertebrates^{1,6–8}. The *Notch* pathway has a key role in vertebrate segmentation (somitogenesis) but is not involved in *Drosophila* body segmentation^{1,7,9}. Here we show that *Notch* and *Delta* genes are involved in segmentation of another arthropod, the spider *Cupiennius salei*. Expression patterns of *Notch* and *Delta*, coupled with RNA interference experiments, identify many similarities between spider segmentation and vertebrate somitogenesis. Our data indicate that formation of the segments in arthropods and vertebrates may have shared a genetic programme in a common ancestor and that parts of this programme have been lost in particular descendant lineages.

Segmented body plans are found in different animal groups, such

as arthropods, vertebrates and annelids, that are not closely related in current phylogenies¹⁰. The complexity of generating a segmented body plan might argue for a common origin of segmentation and a common genetic programme^{1,2}. In the past two decades, however, genetic analyses in the fruitfly *Drosophila*^{3,4} and in various vertebrates such as mouse, chick and zebrafish^{7–9,11–14} suggest that fundamentally different mechanisms and gene networks are involved in arthropod and vertebrate segmentation. *Drosophila* segments are generated by a successive spatial refinement along the anterior–posterior axis under the control of a hierarchical cascade of transcription factors^{3,4}. By contrast, the formation of vertebrate somites involves a molecular oscillator, known as the segmentation clock, that drives the periodic expression of genes in the presomitic mesoderm^{7,8,13,14}. Each wave of oscillatory gene expression results in the formation of one somite. *Notch* and *Delta* genes are crucial components of this vertebrate segmentation clock^{8,15}.

One *Notch* gene and two *Delta* genes (*Delta-1* and *Delta-2*) have been identified in the spider *C. salei*¹⁶. Here we have analysed their role in segmentation. In the spider, segments are sequentially generated from a posterior growth zone¹⁷. The *Notch* gene is expressed in this growth zone and resolves into segmental expression just before the segments form (Fig. 1a, e). There is a stronger accumulation of *Notch* transcripts at the posterior border of the newly formed stripes. *Notch* is expressed in the newly formed segments in the same register as the *engrailed* gene, which defines the parasegment boundary (ref. 6 and data not shown). The segmental expression of *Notch* fades in more anterior segments.

The spider *Delta-1* gene is expressed in a dynamic pattern in the growth zone (Fig. 1c, d). Initially *Delta-1* is expressed in a posterior

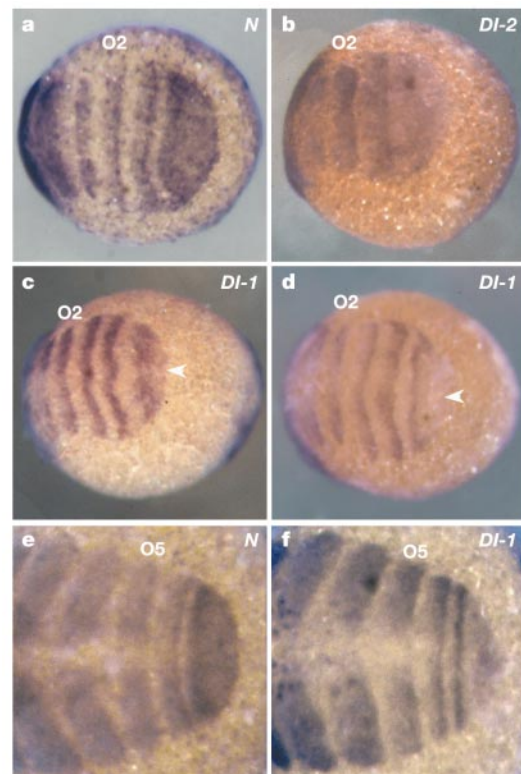


Figure 1 Expression of *Notch*, *Delta-1* and *Delta-2* genes during segmentation in the spider *C. salei*. **a–d**, *In situ* hybridization shows the expression of *Notch* (**a**), *Delta-2* (**b**) and *Delta-1* (**c**, **d**) in the posterior growth zone of embryos. The stripe corresponding to the just forming second opisthosomal segment is labelled O2. Arrowheads in **c** and **d** point to the posterior end of the embryo proper. **e**, **f**, Posterior end of slightly older embryos stained for *Notch* (**e**) or *Delta-1* (**f**). The most recently formed segment, the fifth opisthosomal, is labelled O5. In all embryos anterior is to the left.

domain (Fig. 1c), which resolves into a stripe while the posterior end becomes clear of *Delta-1* transcripts (Fig. 1d). *Delta-1* is then expressed again at the posterior end and a new stripe of *Delta-1* expression forms. In this way, the stripes of *Delta-1* expression form sequentially from the posterior end of the growth zone. This aspect of *Delta-1* expression resembles the expression pattern of the spider orthologues of the *Drosophila* pair rule genes *hairy*, *even-skipped* and *runt*⁵. It remains to be determined whether the dynamic expression of these genes is due to oscillation, as it is in vertebrates. The more anterior *Delta-1* stripes show a strong and sharp posterior border (Fig. 1f). Again, *engrailed* is expressed in the same register as the *Delta-1* stripes (data not shown). The spider *Delta-2* gene is expressed in the posterior growth zone and forms stripes towards anterior (Fig. 1b). But this expression is rather weak and the borders of the stripes are fuzzy.

To analyse the functional role of *Notch* and *Delta* genes in spider segmentation, we used an approach based on RNA-mediated interference (RNAi)¹⁸. Embryos injected with dsRNA targeted at *Notch*, *Delta-1* or *Delta-2* genes showed severe defects in segment patterning and formation of the segment borders (Fig. 2 and Table 1). The phenotypes for the three genes seem to be identical. Patterning defects include malformation of segments: the size, shape and width of the segments are affected and vary. In addition, the segment borders are irregular and not well defined, as can be seen by the disturbed stripes of *engrailed* gene expression, which no longer form straight sharp borders (Fig. 2e–l) as compared with their straight sharp borders in control embryos (ref. 6 and Fig. 2a–d). An additional defect found in the RNAi embryos is the apparent larger and malformed posterior growth zone; this seems to be the result of patterning or morphogenetic defects rather than additional cell divisions, as there are no detectable differences in cell proliferation between RNAi and control embryos (see Supplementary Information).

The spider basic helix–loop–helix (bHLH) gene *hairy* (*Cs-h*) is

Table 1 Effect of *Notch*, *Delta-1* and *Delta-2* RNAi in spider embryos

	Total (n)	Segmentation phenotype	No effect	Unspecific effects
No injection	89	0 (0%)	76 (85%)	13 (15%)
GFP dsRNA	50	0 (0%)	44 (88%)	6 (12%)
<i>Notch</i> dsRNA	160	101 (63%)	38 (24%)	21 (13%)
<i>Delta-1</i> dsRNA	148	89 (60%)	37 (25%)	22 (15%)
<i>Delta-2</i> dsRNA	156	89 (57%)	48 (31%)	19 (12%)

Shown are the number (and percentages) of embryos that show segmentation defects after the injection of dsRNA¹⁸. Control embryos were not injected ('no injection') or were injected with dsRNA targeted to the jellyfish GFP gene encoding green fluorescent protein.

dynamically expressed in stripes in the growth zone of the embryo (ref. 5 and Fig. 3a) in a pattern comparable to that of *Delta-1* (Fig. 1c, d). In a single-colour double *in situ* hybridization assay for both *hairy* and *Delta-1*, there is no more staining than in a single staining assay for *hairy*, implying that the staining for *Delta-1* expression must be covered by the staining for *hairy* in this *in situ* assay (data not shown). *Delta-1* is thus expressed in a subset of *hairy*-expressing cells during the dynamic phase in the growth zone. After the spider *Notch* or *Delta* genes are silenced by RNAi, the *hairy*-expressing cells are no longer organized in stripes but are scattered throughout the growth zone (Fig. 3b, c). Thus, *Notch* signalling is required to organize the expression of *hairy* in stripes in the growth zone of the spider.

The expression of vertebrate *hairy/Enhancer of split* (*E(spl)*)-type bHLH genes, such as zebrafish *her1* and *her7*, mouse *Hes1*, *Hes5* and *Hes7*, and chicken *c-hairy*, in stripes in the presomitic mesoderm also depends on *Notch*, and the expression of these bHLH genes in stripes is disturbed in embryos deficient for *Notch* signalling^{8,11–14,19–21}. By contrast, the *Notch* pathway does not regulate *hairy* in *Drosophila* body segmentation, although it does in eye development²² and it acts on other bHLH genes, such as the *E(spl)* genes, during neurogenesis²³. The disturbance of the *hairy* expression pattern in *Notch* and *Delta* RNAi embryos suggests that these genes are incorporated into one genetic network for body segmen-

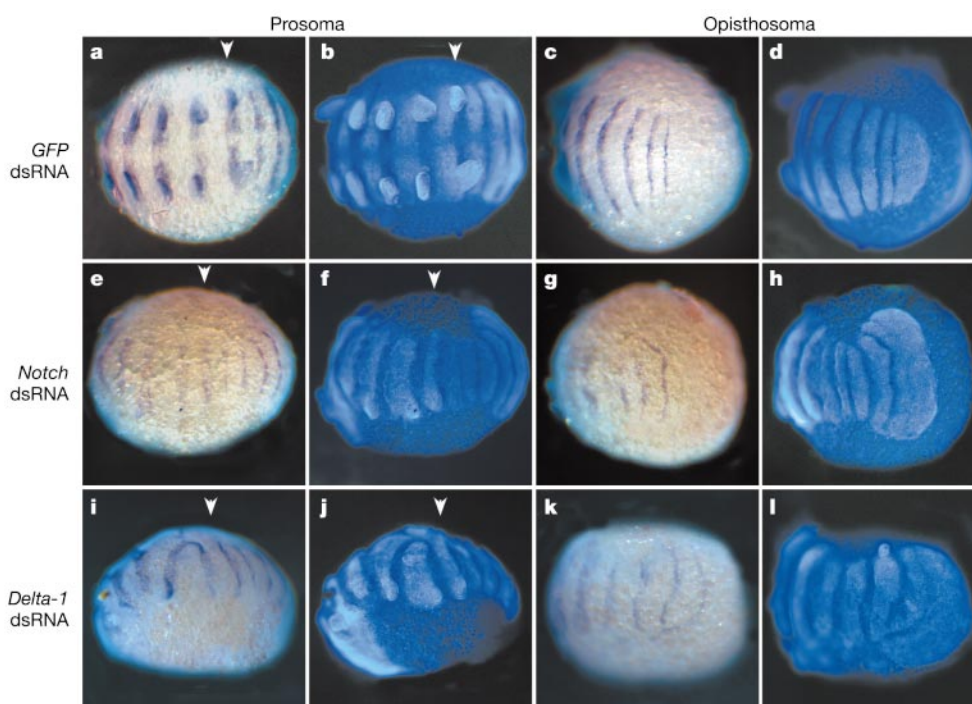


Figure 2 Segmentation defects in *Notch* and *Delta* RNAi embryos. Embryos were injected with dsRNA targeted to GFP (control) (a–d), *Notch* (e–h) and *Delta-1* (i–l). Embryos were stained for the segmental marker *engrailed*⁶. For each embryo the bright field image is shown (a, c, e, g, i, k) together with the epi-fluorescence image of 4',6-diamidino-2-

phenylindole dihydrochloride (DAPI)-stained DNA (b, d, f, h, j, l), respectively. The left images show the prosomal domain of the germ band, the right images show the posterior end of the opisthosoma. Arrowhead marks the fourth walking leg. In all embryos anterior is to the left.

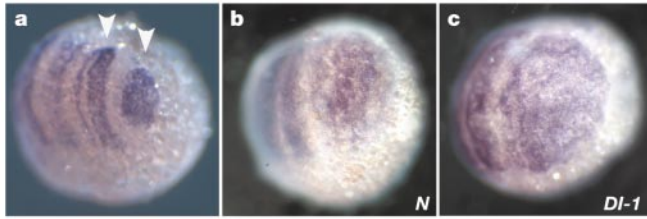


Figure 3 The expression of the *hairy* gene is disturbed in embryos deficient for *Notch* or *Delta*. **a**, In untreated embryos, *hairy* is expressed dynamically in stripes in the growth zone⁶; arrowheads point to the stripes of *hairy* expression in the posterior growth zone. **b**, **c**, The organization of *hairy*-expressing cells in stripes is disturbed after RNAi of *Notch* (**b**) and *Delta-1* (**c**). The *hairy*-expressing cells are scattered throughout the growth zone. In all embryos anterior is to the left.

tation in the spider, similar to the network in vertebrates.

Our data provide evidence for the involvement of the *Notch* signalling pathway in arthropod segmentation and show that *Notch* signalling is required to pattern the segments and to obtain straight sharp segment borders in the spider. By contrast, *Notch* signalling is not known to be required in body segmentation of *Drosophila*. In *Drosophila*, the segments form simultaneously at the syncytial blastoderm stage, when the embryo is not yet cellularized and so cannot rely on cell-to-cell signalling such as that mediated by *Notch*. In the more ancestral mode of arthropod segmentation, however, segments form sequentially from a cellularized posterior growth zone, as in the spider^{17,24}.

Vertebrate somites also form sequentially in the presomitic mesoderm. In vertebrate somitogenesis, *Notch* signalling is required during two phases: first in the segmentation clock that acts in the presomitic mesoderm, and subsequently in establishing the somite borders^{7,11,14,15}. Vertebrate segmentation is a patterning of the mesoderm, whereas the *Drosophila* segmentation gene cascade patterns the ectoderm. The origin and the formation of the mesoderm in spiders is not resolved²⁵; therefore, it is not clear whether *Notch* signalling in the spider segmentation patterns the mesoderm, the ectoderm, or both. The involvement of *Notch* signalling in segmentation of both spiders and vertebrates raises the question of whether the use of *Notch* signalling in spider and vertebrate segmentation is the result of either a common origin or a convergent recruitment of the *Notch* pathway into segmentation. Although *Notch* signalling is not known to be required for body segmentation in *Drosophila*¹, it has a crucial role in other processes involving boundary formation in the eye, the wing and the legs^{26–28}.

There is also evidence for involvement of *Notch* signalling downstream of the segmentation machinery in insects. The *fringe* gene, which encodes a modulator of *Notch* signalling, is expressed in segmental stripes after the formation of the segments in the grasshopper²⁹, a more basal insect, whereas *Serrate*, which encodes a *Notch* ligand, is activated in stripes ventrally in the segments and is required to promote denticle diversity in *Drosophila*³⁰. Thus, involvement of *Notch* signalling in segmentation may have been ancestral for arthropods, but this involvement has been reduced in the lineage leading to *Drosophila* and might be replaced by other genes. Consequently, the observed similarities between spider and vertebrates suggest that the formation of the segments in arthropods and vertebrates may have shared a genetic programme involving *Notch* signalling in a common ancestor and that parts of this programme have been lost in particular lineages, such as the one leading to the higher insects. □

Methods

Spiders

Colonies of the Central American wandering spider *C. salei* Keys (Chelicerata, Araneida, Ctenidae) are maintained in Cologne. We obtained embryos as described⁶.

In situ hybridization and RNAi

Whole-mount *in situ* hybridization with modifications for spider embryos was done as described⁶. RNAi was done as described¹⁸.

Received 18 February; accepted 14 April 2003; doi:10.1038/nature01682.

- Davis, G. K. & Patel, N. H. The origin and evolution of segmentation. *Trends Genet.* **15**, M68–M72 (1999).
- De Robertis, E. M. Evolutionary biology. The ancestry of segmentation. *Nature* **387**, 25–26 (1997).
- St Johnston, D. & Nüsslein-Volhard, C. The origin of pattern and polarity in the *Drosophila* embryo. *Cell* **68**, 201–219 (1992).
- Pankratz, M. & Jäckle, H. in *Development of Drosophila melanogaster* (eds Bate, M. & Martinez-Arias, A.) 467–516 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1993).
- Damen, W. G. M., Weller, M. & Tautz, D. The expression patterns of *hairy*, *even-skipped*, and *run* in the spider *Cupiennius salei* imply that these genes were segmentation genes in a basal arthropod. *Proc. Natl Acad. Sci. USA* **97**, 4515–4519 (2000).
- Damen, W. G. M. Parasegmental organization of the spider embryo implies that the parasegment is an evolutionary conserved entity in arthropod embryogenesis. *Development* **129**, 1239–1250 (2002).
- Saga, Y. & Takeda, H. The making of a somite: molecular events in vertebrate segmentation. *Nature Rev. Genet.* **2**, 835–845 (2001).
- Pourquie, O. Notch around the clock. *Curr. Opin. Genet. Dev.* **9**, 559–565 (1999).
- Conlon, R. A., Reaume, A. G. & Rossant, J. Notch1 is required for the coordinate segmentation of the somites. *Development* **121**, 1533–1545 (1995).
- Aguinaldo, A. M. *et al.* Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature* **387**, 489–493 (1997).
- Holley, S. A., Julich, D., Rauch, G. J., Geisler, R. & Nüsslein-Volhard, C. *her1* and the *notch* pathway function within the oscillator mechanism that regulates zebrafish somitogenesis. *Development* **129**, 1175–1183 (2002).
- Holley, S. A., Geisler, R. & Nüsslein-Volhard, C. Control of *her1* expression during zebrafish somitogenesis by a *delta*-dependent oscillator and an independent wave-front activity. *Genes Dev.* **14**, 1678–1690 (2000).
- Palmeirim, I., Henrique, D., Ish-Horowicz, D. & Pourquie, O. Avian *hairy* gene expression identifies a molecular clock linked to vertebrate segmentation and somitogenesis. *Cell* **91**, 639–648 (1997).
- Jouve, C. *et al.* Notch signalling is required for cyclic expression of the *hairy*-like gene *HES1* in the presomitic mesoderm. *Development* **127**, 1421–1429 (2000).
- Jiang, Y. J. *et al.* Notch signalling and the synchronization of the somite segmentation clock. *Nature* **408**, 475–479 (2000).
- Stolte, A. Recruitment of cell groups through Delta/Notch signalling during spider neurogenesis. *Development* **129**, 5339–5348 (2002).
- Seitz, K.-A. Normale Entwicklung des arachniden Embryos *Cupiennius salei* Keyserling und seine Regulationsbefähigung nach Röntgenbestrahlungen. *Zool. Jb. Anat. Bd.* **83**, 327–447 (1966).
- Schoppmeier, M. & Damen, W. G. M. Double-stranded RNA-interference (RNAi) in the spider *Cupiennius salei*: the role of *Distal-less* is evolutionarily conserved in arthropod appendage formation. *Dev. Genes Evol.* **211**, 76–82 (2001).
- Oates, A. C. & Ho, R. J. *Hairy/E(spl)*-related (*Her*) genes are central components of the segmentation oscillator and display redundancy with the Delta/Notch signaling pathway in the formation of anterior segmental boundaries in the zebrafish. *Development* **129**, 2929–2946 (2002).
- Bessho, Y. *et al.* Dynamic expression and essential functions of *Hes7* in somite segmentation. *Genes Dev.* **15**, 2642–2647 (2001).
- Dale, J. K. *et al.* Periodic Notch inhibition by Lunatic Fringe underlies the chick segmentation clock. *Nature* **421**, 275–278 (2003).
- Baonza, A. & Freeman, M. Notch signalling and the initiation of neural development in the *Drosophila* eye. *Development* **128**, 3889–3998 (2001).
- Jennings, B., Preiss, A., Delidakis, C. & Bray, S. The Notch signalling pathway is required for Enhancer of split bHLH protein expression during neurogenesis in the *Drosophila* embryo. *Development* **120**, 3537–3548 (1994).
- Davis, G. K. & Patel, N. H. Short, long and beyond: molecular and embryological approaches to insect segmentation. *Annu. Rev. Entomol.* **47**, 669–699 (2002).
- Anderson, D. T. *Embryology and Phylogeny in Annelids and Arthropods* 365–451 (Pergamon, Oxford, 1973).
- Papayannopoulos, V., Tomlinson, A., Panin, V. M., Rauskolb, C. & Irvine, K. Dorsal-ventral signaling in the *Drosophila* eye. *Science* **281**, 2031–2034 (1998).
- Brook, W. J., Diaz-Benjumea, F. J. & Cohen, S. M. Organizing spatial pattern in limb development. *Annu. Rev. Cell Dev. Biol.* **12**, 161–180 (1996).
- de Celis, J. F., Tyler, D. M., de Celis, J. & Bray, S. J. Notch signalling mediates segmentation of the *Drosophila* leg. *Development* **125**, 4617–4626 (1998).
- Dearden, P. & Akam, M. A role for *Fringe* in segment morphogenesis but not segment formation in the grasshopper *Schistocerca gregaria*. *Dev. Genes Evol.* **210**, 329–336 (2000).
- Wiellette, E. L. & McGinnis, W. Hox genes differentially regulate *Serrate* to generate segment-specific structures. *Development* **126**, 1985–1995 (1999).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank D. Tautz for support and discussion; S. Roth for comments on the manuscript; T. Klein for discussion; A. Pozhitkov for help with the statistic analysis; and R. Janßen for care of the spider culture. This work was partially supported by the Deutsche Forschungsgemeinschaft.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to W.D. (damen@uni-koeln.de).

Somatosensory basis of speech production

Stéphanie Tremblay*, Douglas M. Shiller* & David J. Ostry*†

* Department of Psychology, McGill University, Montreal, Quebec H3A 1B1, Canada

† Haskins Laboratories, New Haven, Connecticut 06511-6695, USA

The hypothesis that speech goals are defined acoustically and maintained by auditory feedback is a central idea in speech production research^{1–6}. An alternative proposal is that speech production is organized in terms of control signals that subserve movements and associated vocal-tract configurations^{7–9}. Indeed, the capacity for intelligible speech by deaf speakers suggests that somatosensory inputs related to movement play a role in speech production—but studies that might have documented a somatosensory component have been equivocal. For example, mechanical perturbations that have altered somatosensory feedback have simultaneously altered acoustics^{10–14}. Hence, any adaptation observed under these conditions may have been a consequence of acoustic change. Here we show that somatosensory information on its own is fundamental to the achievement of speech movements. This demonstration involves a dissociation of somatosensory and auditory feedback during speech production. Over time, subjects correct for the effects of a complex mechanical load that alters jaw movements (and hence somatosensory feedback), but which has no measurable or perceptible effect on acoustic output. The findings indicate that the positions of speech articulators and associated somatosensory inputs constitute a goal of speech movements that is wholly separate from the sounds produced.

We have adapted a technique used in studies of limb motor control to apply velocity dependent mechanical perturbations to the jaw. The perturbation was designed to be of sufficient strength to alter systematically the motion path of the jaw, and hence somatosensory feedback, without affecting the associated acoustic output. As in work on limb movement, adaptation to an artificial mechanical force field indicates the adjustment of control signals to take account of loads on the basis of sensory input^{15–19}.

We have altered somatosensory feedback in three different tasks—one involving normal vocalized speech, another during ‘silent speech’ (speech without vocalization), and a third that involves a non-speech jaw movement that is matched in amplitude and duration to that observed in speech. In the vocalized speech condition, subjects were required to repeatedly produce the utterance *siat* (pronounced ‘see-at’) at a subject-chosen rate and volume.

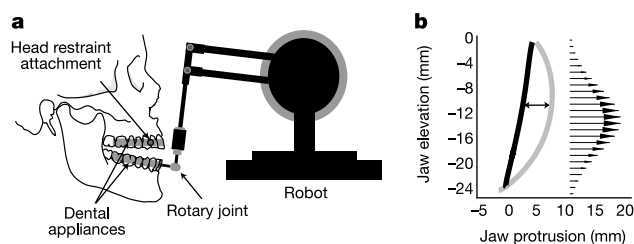


Figure 1 Experimental set-up and representative data. **a**, Diagram showing subject attached to the robotic device. **b**, Jaw opening movement with the force field off (black) and on initial exposure to the field (grey). Vectors depict the magnitude and direction of force applied by the robot over the course of the movement. The double-headed arrow shows the maximum horizontal deviation between null-field and force-field movements that served as a performance index.

This condition was tested to assess the extent to which adaptation to a somatosensory perturbation might occur in the presence of unaltered acoustic feedback. The silent speech condition explicitly removed auditory feedback, and hence examined the ability of subjects to adapt in the total absence of auditory input. Subjects in this group were asked to articulate the utterance *siat* without producing any sound. The non-speech condition addressed the issue of whether adaptation would occur in a cyclical jaw movement task that is matched only in amplitude and duration to that observed in speech. There was no reference whatsoever to speech production in the description of the task for the non-speech group.

A robotic device was connected to the mandibular teeth, and was used to deliver mechanical perturbations to the jaw (Fig. 1a). Sagittal plane forces were applied along a horizontal axis (parallel to the occlusal plane), in the direction of jaw protrusion. The forces were proportional to the instantaneous vertical velocity of the jaw (measured at the incisors) such that the magnitude of the perturbation increased with the velocity of movement (Fig. 1b). Performance was quantified for each subject by measuring on a trial-by-trial basis the maximum horizontal distance between the movement path under force-field conditions and the average movement path with the field off (null field). A decrease in the horizontal distance over trials reflects sensorimotor adaptation in which the effect of the force field is reduced.

Analyses of kinematic data revealed a systematic pattern of force-field adaptation in speech production. Figure 2 illustrates movements for individual subjects in the vocalized speech condition (Fig. 2a), the silent speech condition (Fig. 2b) and the non-speech condition (Fig. 2c). A baseline phase of 20 null-field trials provided a reference movement path under unperturbed conditions (black lines). As shown in blue, the jaw path deviated in the direction of protrusion with the introduction of the force field. Following training, adaptation (shown in red) was observed in the vocalized speech and the silent speech conditions, but not in the non-speech condition. The green paths illustrate an after-effect in which the jaw was retracted in comparison to the baseline following the unexpected removal of the force field (vocalized speech and silent speech conditions).

Figure 3 gives mean values of maximum horizontal deviation in each of these conditions on a per-subject basis. It can be seen that the force field had a similar initial effect for subjects in all experimental conditions: compared to the baseline, movements at the start of training in the force field deviated significantly in the protrusion direction ($P < 0.001$ for all subjects). Although there

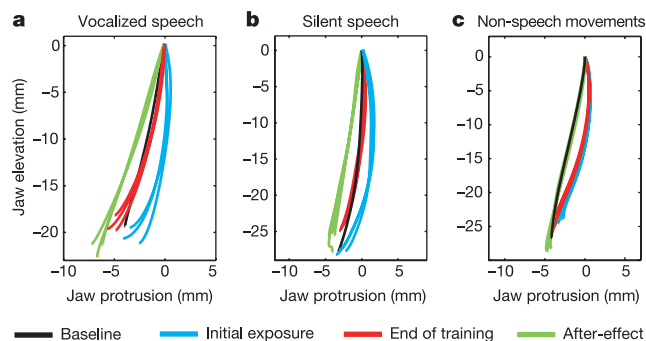


Figure 2 Sagittal plane jaw motion paths. Data were acquired during the baseline condition (black trace), on initial exposure to the force field (blue), at the end of training (red), and following unexpected removal of the field (green). The figure shows individual trials for single subjects. **a**, During vocalized speech, adaptation to the force field and a subsequent after-effect are observed. **b**, During silent speech, the pattern of adaptation and after-effect observed in vocalized speech are unaltered by removal of acoustic feedback. **c**, Matched non-speech movements show neither adaptation nor an after-effect.

were individual differences in the initial deviation caused by the field, there was no overall difference in the magnitude of the perturbation between the groups ($P > 0.05$). Somatosensory feedback was thus initially altered in a comparable way in each of the conditions tested. At the end of training (shown in red), adaptation was observed as indicated by a significant reduction in deviation from baseline for all but two subjects (S7 and S8) in the vocalized speech group ($P < 0.01$), and for all subjects in the silent speech group ($P < 0.001$) (Fig. 3a, b). In contrast, adaptation was not observed in the non-speech group: for two subjects, movements at the end of training did not differ reliably from those at the beginning ($P > 0.05$), for the other two subjects there was an increase in the deviation (Fig. 3c, left side). A motion-dependent after-effect (shown in green) following the unexpected removal of the force field illustrates that adjustments to offset the effects of the force field were generally as large in magnitude as the initial deviation for the vocalized speech and the silent speech groups (Fig. 3a, b). None of the four subjects in the non-speech condition showed an after-effect in the direction of retraction relative to baseline (Fig. 3c).

Two subjects in the non-speech group were tested over an extended period of time to see whether the absence of adaptation in that condition arose simply as a consequence of limited practice. Both were tested under force-field conditions for four days in a row (2,100 trials in total for each subject). The right side of Fig. 3c shows results of the final day of training, labelled 2B and 3B. Neither of the

subjects showed an adaptation or an after-effect. This indicates that matching jaw movements on duration and amplitude alone is insufficient for the achievement of the adaptation observed in the speech groups.

As a further control, we tested a variant of the non-speech condition in which subjects were required to produce discrete jaw lowering movements from an initial position with the mouth closed to a remembered target location. We reasoned that discrete jaw lowering movements might have absolute spatial goals more comparable to those observed in speech. The opening movements were equal in amplitude and duration to the opening movements in the vocalized speech condition. Subjects were instructed to briefly hold the target position and then to return slowly to the starting position. Subjects were trained in the same force field as in the previous conditions. A pattern similar to that observed in cyclical non-speech movements was obtained (Fig. 3d). For all subjects, initial exposure to the force field resulted in a reliable deviation from baseline conditions ($P < 0.01$). Training did not produce a reduction in the deviation ($P > 0.05$). The sudden removal of the force field did not result in an after-effect comparable to that observed in Fig. 3a, b ($P > 0.05$).

To verify that the perturbations produced by the force field did not systematically alter the speech acoustics, the first and second formant frequencies were examined during the transition between *i* and *a* in the test utterance *siat*. Figure 4a gives an example of the acoustic spectrogram for a single utterance. The spectrogram depicts F1 and F2 frequencies during the voiced portion of the utterance. The white rectangle highlights the transition between vowels. Figure 4b shows transitions in formant frequencies between vowels (time-normalized) at different phases of the experiment for a single subject. It may be seen that the formant frequency transitions were similar throughout the experiment. Statistical tests were carried out across subjects in order to compare the acoustic signals in the baseline, the first and the last 20% of training, and in the final null-field trials in which the load to the jaw had been unexpectedly removed. Differences in average F1 and F2 frequency were assessed at the midpoint of the vowel transition as well as at its start and end

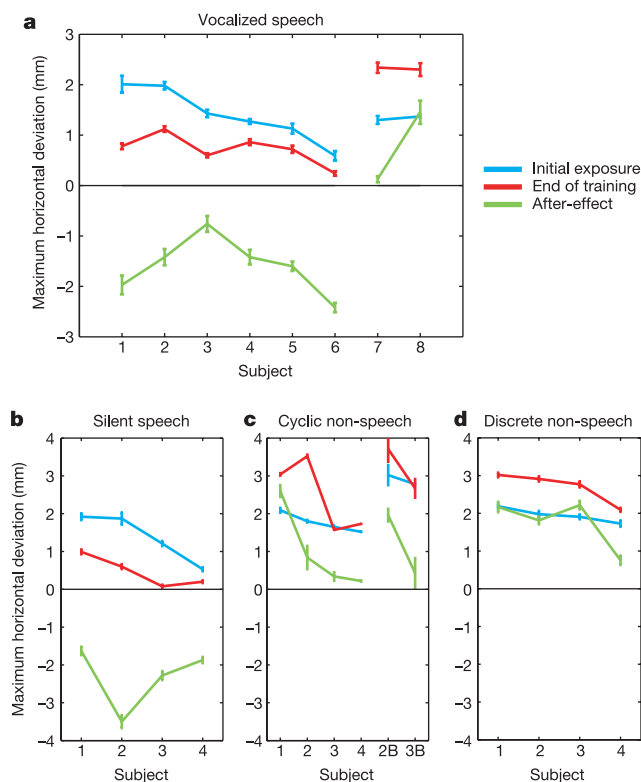


Figure 3 Jaw position during force-field adaptation shown on a per-subject basis. Average values of maximum horizontal deviation are presented for initial exposure to the force field (blue), at the end of the training (red), and following the unexpected removal of the field (green). Connecting lines are provided for visualization purposes only, and do not imply correlations across subjects. **a**, In vocalized speech, six out of eight subjects showed adaptation (reduction of deviation and an after-effect). **b**, All subjects in the silent speech condition showed adaptation. **c**, None of the subjects in the cyclical non-speech condition showed either a reduction in deviation or an after-effect (S1–S4). S2 and S3 were tested for an extended period of time and showed no adaptation (2B and 3B). **d**, Subjects in the discrete non-speech condition showed a pattern similar to those in the cyclical non-speech condition.

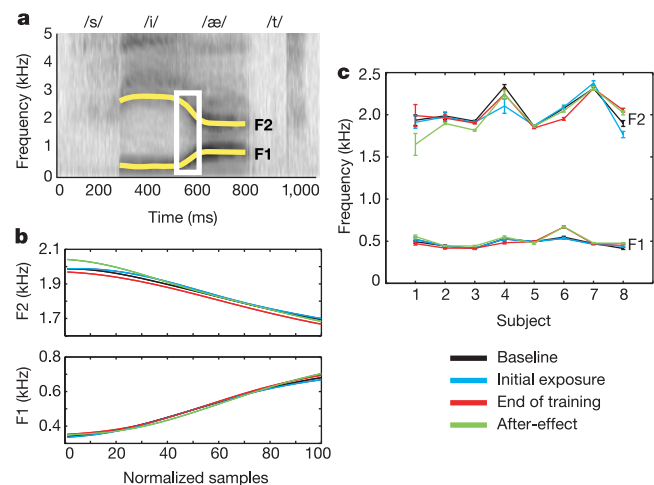


Figure 4 Acoustic data. **a**, Spectrogram of the acoustic signal during a single repetition of the utterance *siat*. The dark bands show the frequency composition of acoustical energy. F1 and F2 formant tracks are indicated by yellow lines. The white rectangle indicates the region of acoustical transition between vowels *i* and *a*. **b**, First and second formant frequencies during the transition from *i* to *a*: baseline (black), initial exposure (blue), end of training (red), and following unexpected removal of the field (green). Formant frequency trajectories for a single subject are shown. The curves give average values for individual blocks. **c**, No systematic acoustic effect is observed when F1 and F2 frequencies are examined on a per-subject basis.

(25% of the peak velocity of the transition). These points were chosen to ensure that the velocity of the vertical movement was sufficiently high to induce a significant mechanical perturbation due to the field. No significant difference was found between the four phases of the experiment at the three measurement points in F1 ($P > 0.05$) or in F2 ($P > 0.05$). Thus, the perturbation did not measurably alter the acoustic signals.

Figure 4c shows the pattern of acoustic effects for each subject separately. The individual data points give the formant frequencies at the peak velocity of the transition between vowels. It may be seen that there is no systematic pattern of acoustic changes associated with the force field. Significant differences in formant frequencies between the four phases of the experiment were observed in a small number of subjects (subject 6 for F1 frequency, and subjects 6 and 8 for F2, $P < 0.01$). In all but one case (F2 frequency for subject 8), values for baseline and initial exposure to the force field were comparable. An examination of formant frequencies on a per-subject basis was also undertaken at the start and end of the acoustic transition between vowels. As in the case of the peak velocity measure, no systematic acoustic effect was observed.

An additional acoustic analysis was carried out to examine the possibility of an acoustic effect on the very first trial in which the force field was applied. No difference was observed in F1 frequency ($P > 0.05$) or in F2 frequency ($P > 0.05$) between the first force-field trial and the remaining three experimental phases (baseline, end of training and after effect). These results further support the idea that the mechanical perturbation created by the force field was not large enough to alter measurably the acoustic signals.

Possible perceptible differences in the acoustics due to the presence of the force field were assessed using a perceptual discrimination task adapted from ref. 20. The results of these perceptual tests showed no ability of listeners to distinguish utterances produced in the force-field condition from those in the baseline condition. The average proportion of correct identification of utterances produced in the force-field condition was 0.54 (99% confidence interval, CI, ± 0.064). These tests were repeated by having the subjects of the vocalized speech condition make perceptual discrimination judgments on their own utterances. For these subjects, the average proportion correct was 0.49 (99% CI ± 0.063). There was thus no perceptible auditory cue that would distinguish utterances produced in force-field and null-field conditions.

In summary, we have examined the role of somatosensory information in speech production by applying velocity-dependent mechanical loads to the jaw that altered the movement path, and hence somatosensory feedback, without affecting the speech acoustics. We found that when speech acoustics were unaltered, or even absent altogether, subjects nevertheless adapted to the perturbation such that the motion path of the jaw approached that observed in the absence of load. The observation that adaptation in jaw movements occurs when speech acoustics are unaltered by the perturbation indicates that somatosensory information on its own is a principal component of the speech target. Changes to somatosensory input result in modifications to speech movements that restore the normal path of the jaw even when the acoustic goal is achieved. The similarity of movements following adaptation to those observed in the absence of load suggests that a precise pattern of somatosensory feedback related to the entire course of the movement is used to update the control signals underlying jaw motion in speech. Moreover, the non-speech conditions demonstrate that adaptation is not an inevitable consequence of training in a force field, that is, it is neither due to reflexes nor a consequence of the active force-generating properties of jaw muscles. Our findings indicate that in speech, a somatosensory goal is pursued independent of the acoustics. $\square$

Methods

Experimental procedures

Mechanical perturbations were delivered to the jaw using a robotic device (Phantom 1.0, Sensable Technologies). The coupling between the robot and the jaw involved (1) an acrylic and metal dental appliance that was glued to the buccal surface of the teeth, and (2) a magnesium and titanium rotary connector that permitted motion of the jaw in all six translational and rotational degrees of freedom. The head was immobilized by connecting a second dental appliance, which was attached to the maxillary teeth, to a rigid metal frame that consisted of two articulated metal arms, one on each side of the head, that were locked in place during data collection.

The force vector, $\mathbf{f}$, produced by the robot depended on the velocity vector of the jaw at the incisors, $\mathbf{v}$, according to the following linear equation: $\mathbf{f} = B \cdot \mathbf{v}$, where B is a constant matrix representing viscosity in N s m^{-1} . Specifically, we used a force field defined by:

$$B = \begin{bmatrix} B_{xx} & B_{xy} \\ B_{yx} & B_{yy} \end{bmatrix} = \begin{bmatrix} 0 & 20 \\ 0 & 0 \end{bmatrix}$$

where B_{xx} represents force in the horizontal direction in proportion to horizontal velocity, and B_{xy} represents horizontal force in proportion to vertical velocity. Peak forces ranged from 4 to 5 N.

Twenty subjects were divided into two groups, a voicing group (8 subjects) and a non-voicing group (12 subjects). The non-voicing group was further separated into three conditions: a silent speech condition (4 subjects), a non-speech cyclical movement condition (4 subjects), and a non-speech discrete movement condition (4 subjects). Subjects in all conditions were instructed to keep movement amplitude and duration constant. The experimenter monitored a real-time display of movement parameters, and provided verbal feedback when amplitude or duration deviated from their initial values by more than $\sim 20\%$. Subjects in the voicing group were also instructed to keep volume constant by monitoring values on a sound pressure level meter.

The experiment began with a familiarization phase with the field off (null field), in which subjects produced 30 repetitions (trials) of the utterance *siat* or the non-speech movement that was to be subsequently tested in the experiment. A baseline phase of 20 further null-field trials provided a reference movement path under unperturbed conditions. This was followed by a field-on training phase of 525 trials, after which the force field was unexpectedly removed and 30 further trials were collected. These final trials under null-field conditions assessed the possible presence of movement after-effects. Data analyses focused on jaw lowering alone, as this phase of movement was associated with potential acoustic effects in the vowel-to-vowel transition (*i-a*) due to the perturbation.

Subject performance was quantified by measuring maximum horizontal distance between perturbed movements and the average baseline path. The analyses were repeated using the horizontal distance between the perturbed movement at maximum lowering velocity and the baseline. The two yielded similar results. Statistical analyses using analysis of variance (ANOVA) were conducted for each subject separately. The analyses compared the maximum deviation from baseline movements in the first 20% of the training trials, the last 20% of the training trials, and the null-field trials following training (after-effect). Pair-wise comparisons of means were carried out using Tukey's method, where appropriate.

Acoustical analyses

The acoustic signal associated with the production of utterances in the vocalized speech condition was analogue low-pass-filtered at 10 kHz and sampled at 22.5 kHz. Acoustic formant tracking was carried out on the transition between the vowels *i* and *a*. For the purpose of subsequent analyses, values for the first and second formants (F1 and F2, respectively) were taken at 25% of the peak velocity of each formant transition (at the beginning and the end of the transition) and also at the point of peak velocity. Note that the acoustics associated with the vowel-to-vowel transitions were of particular interest as the force field depended on movement velocity and hence would have had the greatest effect during the transition between vowels. Statistical tests of potential acoustic effects due to the force field were conducted on a per-subject basis for the first and second formant frequencies separately using ANOVA.

Perceptual task

Six listeners were presented with randomly selected individual utterances recorded from null-field and initial force-field trials. Each test sequence comprised four utterances. The sequences were either of the form AABA or ABAA, where A represents an utterance randomly chosen from null-field trials in the baseline condition and B is an utterance randomly selected from initial force-field trials. The listeners were told that they would hear utterances that were recorded in two different conditions, and were required to indicate whether the second or the third differed from the other three. Listeners were presented with three blocks of 50 such sequences. Each block contained utterances selected from a single speaker.

The test was repeated using four subjects in the vocalized speech condition who were required to make perceptual judgments on their own utterances. In this case, two blocks of 50 sequences composed entirely of the listeners' own productions were used as stimuli. All other aspects of the procedure were identical to those described above.

Received 24 March; accepted 10 April 2003; doi:10.1038/nature01710.

1. Guenther, F. H., Hampson, M. & Johnson, D. A theoretical investigation of reference frames for the planning of speech movements. *Psychol. Rev.* **105**, 611–633 (1998).
2. Guenther, F. H. et al. Articulatory tradeoffs reduce acoustic variability during American English /r/ production. *J. Acoust. Soc. Am.* **105**, 2854–2865 (1999).

3. Houde, J. F. & Jordan, M. I. Sensorimotor adaptation in speech production. *Science* **279**, 1213–1216 (1998).
4. Jones, J. A. & Munhall, K. G. Perceptual calibration of F0 production: Evidence from feedback perturbation. *J. Acoust. Soc. Am.* **108**, 1246–1251 (2000).
5. Perkell, J. S., Matthies, M. L., Svirsky, M. A. & Jordan, M. I. Trading relations between tongue-body raising and lip rounding in production of the vowel /u/: A pilot “motor equivalence” study. *J. Acoust. Soc. Am.* **93**, 2948–2961 (1993).
6. Perkell, J. S. & Nelson, W. L. Variability in production of the vowels /i/ and /a/. *J. Acoust. Soc. Am.* **77**, 1889–1895 (1985).
7. Browman, C. P. & Goldstein, L. Articulatory phonology: An overview. *Phonetica* **49**, 155–180 (1992).
8. Guenther, F. H. Speech sound acquisition, coarticulation and rate effects in a neural network model of speech production. *Psychol. Rev.* **102**, 594–621 (1995).
9. Saltzman, E. L. & Munhall, K. G. A dynamical approach to gestural patterning in speech production. *Ecol. Psychol.* **1**, 333–382 (1989).
10. Hamlet, S. L. & Stone, M. L. Compensatory vowel characteristics resulting from the presence of different types of experimental dental prostheses. *J. Phonet.* **4**, 199–218 (1976).
11. Lindblom, B., Lubker, J. & Gay, T. Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation. *J. Phonet.* **7**, 147–161 (1979).
12. McFarland, D., Baum, S. R. & Chabot, C. Speech compensation to structural modifications of the oral cavity. *J. Acoust. Soc. Am.* **100**, 1093–1104 (1996).
13. McFarland, D. & Baum, S. R. Incomplete compensation to articulatory perturbation. *J. Acoust. Soc. Am.* **97**, 1865–1873 (1995).
14. Savariaux, C., Perrier, P. & Orliaguet, J. P. Compensation strategies for the perturbation of the rounded vowel [u] using a lip tube: A study of the control of space in speech production. *J. Acoust. Soc. Am.* **98**, 2428–2442 (1995).
15. Gandolfo, F., Mussa-Ivaldi, F. A. & Bizzi, E. Motor learning by field approximation. *Proc. Natl Acad. Sci. USA* **93**, 3843–3846 (1996).
16. Goodbody, S. J. & Wolpert, D. M. Temporal and amplitude generalization in motor learning. *J. Neurophysiol.* **79**, 1825–1838 (1998).
17. Lackner, J. R. & Dizio, P. Rapid adaptation to coriolis force perturbations of arm trajectory. *J. Neurophysiol.* **72**, 299–313 (1994).
18. Shadmehr, R. & Mussa-Ivaldi, F. A. Adaptive representation of dynamics during learning of a motor task. *J. Neurosci.* **14**, 3208–3224 (1994).
19. Thoroughman, K. A. & Shadmehr, R. Learning of action through adaptive combination of motor primitives. *Nature* **407**, 742–747 (2000).
20. Bernstein, L. R. & Trahiotis, C. Detection of interaural delay in high-frequency noise. *J. Acoust. Soc. Am.* **71**, 147–152 (1982).

Acknowledgements We thank G. Houle, M. Tiede and C. Dolan for technical support. This work was supported by the National Institute on Deafness and Other Communication Disorders, the Natural Sciences and Engineering Research Council of Canada, and Le Fonds pour la Formation de Chercheurs et l'Aide à la Recherche, Quebec.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.J.O. (ostry@motion.psych.mcgill.ca).

The influence of visual motion on fast reaching movements to a stationary object

David Whitney*, David A. Westwood† & Melvyn A. Goodale*

* CIHR Group on Action and Perception, The Department of Psychology, The University of Western Ontario, London, Ontario N6A 5C2, Canada

† School of Health and Human Performance, Dalhousie University, Halifax, Nova Scotia B3H 3J5, Canada

One of the most important functions of vision is to direct actions to objects¹. However, every time that vision is used to guide an action, retinal motion signals are produced by the movement of the eye and head as the person looks at the object or by the motion of other objects in the scene. To reach for the object accurately, the visuomotor system must separate information about the position of the stationary target from background retinal motion signals—a long-standing problem that is poorly understood^{2–7}. Here we show that the visuomotor system does not distinguish between these two information sources: when observers made fast reaching movements to a briefly presented

stationary target, their hand shifted in a direction consistent with the motion of a distant and unrelated stimulus, a result contrary to most other findings^{8,9}. This can be seen early in the hand's trajectory (~120 ms) and occurs continuously from programming of the movement through to its execution. The visuomotor system might make use of the motion signals arising from eye and head movements to update the positions of targets rapidly and redirect the hand to compensate for body movements.

In the first experiment we investigated the role of distant motion signals in the updating of reaching movements to a stationary target. We briefly presented a stationary object while subjects fixated on a bull's-eye at the centre of a screen (see Fig. 1a and Methods). Subjects hit the position of the flashed object as quickly as possible with their index finger. A vertically drifting pattern was presented on the screen throughout the trial. Initially the pattern drifted in one direction, but, at an unpredictable moment, the direction of the drifting pattern was reversed.

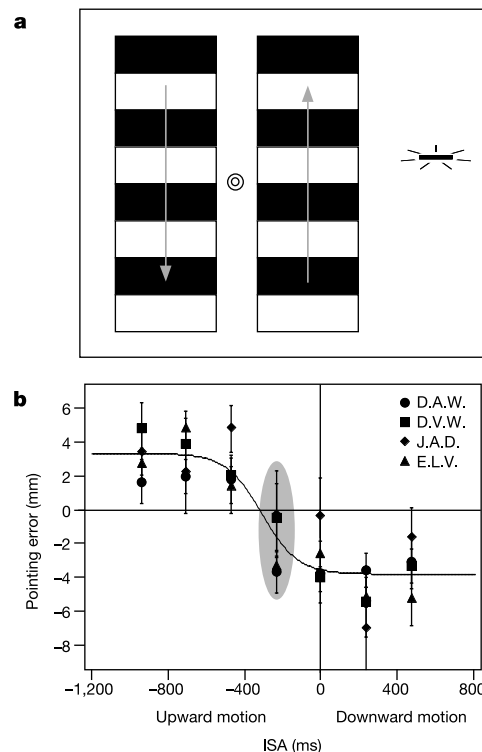


Figure 1 Stimulus and results for the first experiment. **a**, Two patterns were presented that drifted vertically in opposite directions. The two patterns reversed direction after a randomly determined interval. Before or after this reversal, a stationary target was briefly presented near the pattern on the right. The target was stationary, rather than moving, because this allows a measurement of position without the confounding effects of target speed. **b**, Results of the first experiment for four subjects. The abscissa shows the ISA between the flashed target and the motion reversal. (Negative ISA indicates that the target was presented before the motion reversal.) The motion reversal is depicted by the vertical line at 0 ISA. Data have been merged so that the motion nearest the target was upwards, then downwards, as indicated along the abscissa. The ordinate represents the magnitude of the hand's endpoint deviation, which was calculated as the average endpoint of the hand for initially upward motion trials minus initially downward motion trials (see Methods). Each data point shows the time at which the target was physically presented and the hand's endpoint deviation (although the data points show the time at which the targets were presented, reaching movements were completed ~500 ms later). The influence of motion on the endpoint position of the hand varied systematically over time for each subject; the least significant effect was for subject J.A.D. ($F_{(6,77)} = 3.14$, $P = 0.008$). The solid line is a sigmoid fit to the average of the four subjects' data. Results are means $\pm$ s.e.m.

Figure 1b shows that the endpoints of the reaching movements were shifted either upwards or downwards in the direction of the nearby motion. Target flashes presented well before (for example -940 ms) or well after (for example 470 ms) the motion reversal (at 0 ms) produced systematic shifts in the hand's position. Because the target in these cases was presented sufficiently long before or after the motion reversal, the entire reaching movement—both programming and execution—was performed during unidirectional motion (average movement onset and movement duration were 224 and 262 ms, respectively). For this reason it is unclear whether the motion influenced the programming phase or the on-line phase of the reaching movement, or both.

The target flash presented 235 ms before the moving pattern reversed direction addresses this question: because the average reaction time was 220 ms, motion was in one direction during most of the programming phase and in the opposite direction during the movement. This condition produced a markedly reduced shift in the movement endpoint (grey oval in Fig. 1b). The only way in which the shift in the hand's endpoint could be reduced to this extent would be if the contributions of visual motion to programming and on-line control were approximately equal.

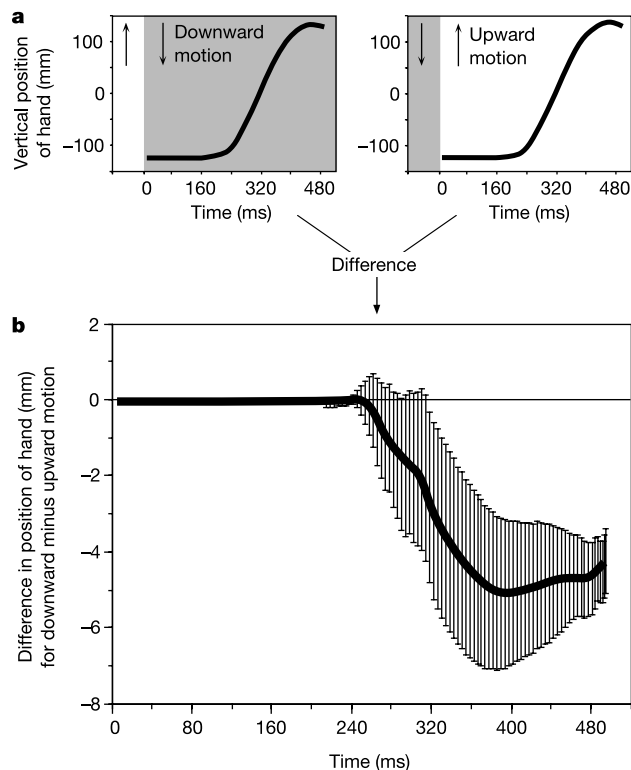


Figure 2 Average hand trajectories from experiment 1 for subject E.L.V., where the target flash occurred coincidentally with the motion reversal. **a**, The two trajectories show both sequences of visual motion (upwards followed by downwards, and vice versa). The grey region in each box represents the presentation of downward visual motion, and the white region indicates upward motion. The abscissa in each graph shows time; the target was presented at 0 ms. The ordinate shows the hand's vertical position (the visual motion on the screen was drifting vertically, so we were interested in the vertical position of the hand). **b**, The hand's trajectory during upward motion in **a** is subtracted from that for downward motion to give the difference in hand position over time. The ordinate shows the shift in the hand's position: negative values indicate that the hand deviated in the direction of visual motion. In this example condition, the target was presented synchronously with the motion reversal (0 ms), so the entire reaching movement was performed during unidirectional visual motion. There is a significant deviation in the trajectory of the hand as a function of time ($F_{(73,584)} = 3.2$, $P < 0.001$). Results are means $\pm$ s.e.m.

Fast reaching movements were subject to on-line control, which is consistent with previous studies in which the target's location is physically altered at the beginning of the reaching movement^{10–14}. Because a continuously moving background was present in this first experiment, the data allow a close examination of how the representation of target location is updated over time and how this representation is used to adjust an ongoing response.

The left-hand side of Fig. 2a shows the average hand trajectory for one subject when there is downward visual motion near the target during the course of the reach (this sample trajectory is taken, in part, from the data point at 0 ms interstimulus asynchrony (ISA) for subject E.L.V. in Fig. 1b; initially the visual motion nearest the target was upwards, but when the target was presented the motion reversed direction, so that the reach was executed during downward motion only). When the trajectory for upward motion (right-hand side of Fig. 2a) is subtracted from this, giving the difference in hand position over time (Fig. 2b), it is clear that the hand continuously shifted in the direction of the nearby motion (increasingly negative values indicate a stronger influence of visual motion on hand position; see Supplementary Information for additional data).

One of the purposes of using a drifting pattern that reversed direction was to reveal the time course of the influence of motion on hand position. Figure 3a shows the average hand trajectories during upward and downward motion in response to the target presented 235 ms before the motion reversed direction (results are for the

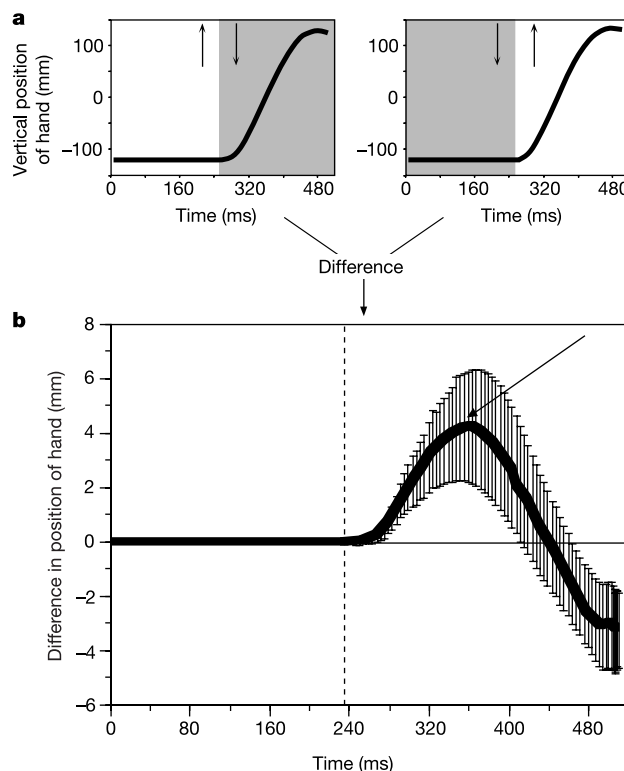


Figure 3 Average hand trajectories from experiment 1 for subject ELV, where the target was presented 235 ms before the motion reversal. **a**, As in Fig. 2, the grey region in each box represents the presentation of downward motion, and the white region represents upward motion. The target was presented at 0 ms. **b**, The trajectories of the hand for the two sequences of motion in **a** are subtracted (upward–downward minus downward–upward). The vertical dashed line represents the motion reversal. The ordinate shows the shift in the hand's position over time. The hand's position began to deviate in the initial direction of motion (increasingly positive values), but eventually the hand shifted in the opposite direction (arrow), which is consistent with the subsequent (reversed) direction of motion. There is a significant deviation in the trajectory of the hand as a function of time ($F_{(80,320)} = 12.96$, $P < 0.001$). Results are means $\pm$ s.e.m.

same subject as in Fig. 2). Figure 3b shows the difference between these curves, which reveals the influence of the motion reversal on the hand's trajectory. In this case, visual motion is in one direction during most of the programming phase and in the opposite direction during the execution of the reaching movement. Interestingly, the trajectory of the hand follows the same pattern: the hand deviates in the direction of the initial motion and subsequently shifts in the opposite direction, mimicking the motion reversal.

Because there is a significant visuomotor delay, some time must pass before visual information can influence the trajectory of the hand. This delay can be estimated from the data in Fig. 3b. Immediately after movement onset (236 ms), the hand begins to deviate in the direction of the initial visual motion. The hand eventually starts to shift in the opposite direction (arrow in Fig. 3b), reflecting the influence of the motion reversal on the hand. The delay between the actual motion reversal (dashed vertical line at 235 ms) and the moment that the data curve reaches a plateau (350 ms) gives a minimum visuomotor delay of ~ 114 ms (see Supplementary Information for an alternative method of calculating the delay). The upper limit on the visuomotor delay is ~ 201 ms, based on the moment (437 ms) at which the average deviation of the hand's position differs significantly from that estimated from the minimum visuomotor delay ($t_{(73)} = 2.05$, $P < 0.05$); this is consistent with the fact that the hand's relative position becomes negative at 437 ms. (Estimates of visuomotor delay for additional

subjects are provided in Supplementary Information). The range of this estimated visuomotor delay (~ 114 – 201 ms) is similar to that for actual changes in target location^{1,10,12,14,15}, indicating that the influence of visual motion on the updating of fast reaching movements occurs on the same time scale as actual changes in target position; that is, motion-generated position reassignment might be equivalent to a shift in the real position of the target¹⁵. This is surprising, because it indicates that information unrelated to the target (extraneous visual motion) might be processed as fast as information specific to the target, such as actual target location.

The present results show that visual motion information can cause shifts in fast reaching movements to the location of a briefly presented, unrelated stationary object. Previous studies have found that goal-directed reaching can, in some circumstances, be influenced by perceptual illusions, indicating that the awareness of a stimulus might determine the behavior^{15–22}. Figure 3b shows that this did not occur in the present experiment. The trajectory of the hand was modified continuously as the direction of visual motion changed. In this particular case, the hand first moved in the direction of the initial motion (for example upwards, after upward visual motion; the first significant deviation of the hand upwards occurred ~ 35 ms after movement onset). However, the target flash is never perceived to be shifted upwards in this situation; the flash always appears either shifted downwards or not shifted at all⁷. The hand initially moved upwards, which is in a direction opposite to that of the perception. If reaching movements depended on awareness of the target's location, the hand should never have been shifted upwards; clearly visual motion influences the representation of target position for fast reaching movements without requiring explicit awareness of the target's position.

One possibility is that the visual motion influenced the perceived speed or position of the hand. To confirm that visibility of the hand is not necessary, we repeated the experiment while visibility of the hand was occluded. The results (see Supplementary Information) were similar to those of the first experiment (Fig. 1b), indicating that the influence of motion on reaching is not due to the visual representation of hand speed or location.

The influence of visual motion on fast reaching movements is greatly reduced when there are significant cues to the target object's position. For example, when the duration of the target is increased sufficiently, the endpoints of the movements are accurate (Fig. 4a). This is an interesting situation, because when the visual information about the location of the target is first used to guide the reaching movement, the visuomotor system has no knowledge of the duration of the target; in other words, at the beginning of the programming phase a brief target flash is identical to a long one, as far as the motor system is concerned. Therefore, for long-duration targets we might expect the trajectory of the hand to deviate in the direction of the visual motion early in the movement, but then to correct itself as the duration of the target increases. This is precisely what we observed (Fig. 4b). During the initial phase of the reach, the surrounding motion signals influenced the position of the hand. However, when there was continued retinal information about target location, the representation of position was recalibrated and the hand's trajectory was updated on-line. The implication is that for any abruptly appearing object, even one that remains visible, there is an influence of visual motion on the hand's early trajectory. The influence of motion on fast reaching movements (revealed in the first experiment) is therefore not restricted to flashed targets that disappear long before the onset of the hand's movement.

Fast reaching movements ultimately depend on a comparison of target and hand position in a common coordinate system^{10,23,24}, a comparison that is likely to be computed only on demand²⁵. However, information about target location must initially be represented in retinotopic coordinates. If this early representation of space were influenced by motion, we would naturally expect

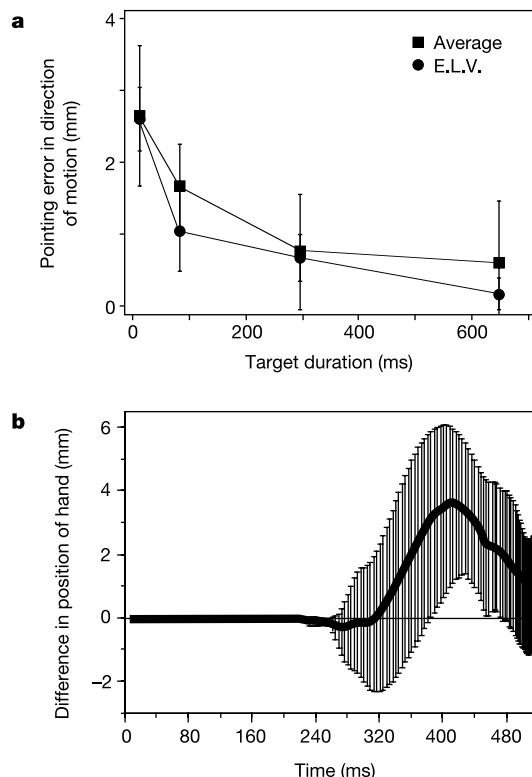


Figure 4 Deviations in reaching movements as a function of target duration. **a**, The endpoints of reaching movements to briefly presented targets were shifted in the direction of visual motion. When the duration of the target exceeded ~ 600 ms, reaching movement endpoints did not exhibit the bias. Data for E.L.V. and the average for nine subjects are shown. **b**, The error in the hand's trajectory is plotted for the 640 ms target duration condition in **a**. See Supplementary Information for additional results. As in Figs 2b and 3b, the difference in the hand's position for the two directions of motion is plotted. Positive values on the ordinate indicate that the hand deviated in the direction of visual motion. There is a significant overall effect of motion on the position of the hand ($t_{(299)} = 4.97$, $P < 0.001$). Results are means $\pm$ s.e.m.

subsequent processing that hinges on this information to manifest a similar distortion. Indeed, motion information is known to influence position coding early in visual processing, even at the level of the retina²⁶ (and probably the cortex^{2,4,5,7,27–29}). It is therefore conceivable that retinal motion signals, even those unrelated to the target, such as the vertically drifting grating in our experiments or the motion across the retina during eye and head movements, might cause continuous updating of target positions for reaching movements. Because the most common source of retinal motion is eye and head movements, an intriguing possibility is that the influence of visual motion on the representation of target position is an adaptive mechanism that allows the hand's trajectory to compensate rapidly for movements of the eye and head with respect to the world.

If visual motion is used to help compensate for ego-motion when executing actions, then we might expect the removal of motion cues to cause increased errors in goal-directed reaching movements. In an experiment, subjects reached to targets while they were passively rotated in a chair (that is, subjects did not have control over their rotation). This form of ego-motion was employed because visual motion might be especially useful for the compensation of passive (unintended) movements of the body. There were three conditions in this experiment. In the first, subjects reached to the remembered locations of targets (light-emitting diodes (LEDs)) while in complete darkness (no visual cues were available). Because this removes not only retinal motion information but also static position cues, a second condition presented intermittent background illumination at ~4 Hz (the background consisted of 25 LEDs that were strobed; see Methods). This effectively eliminated motion information but left static position cues intact. In the third condition—the only condition in which motion information was available—the 25 background LEDs were continuously visible. When no background was visible, endpoints of the reaching movements were biased in the direction of the subject's rotation (subjects overshot the target, underestimating the magnitude of their rotation, which is consistent with previous studies³⁰). There was a significant improvement (decrease) in this bias when retinal motion was available, compared with both the intermittent illumination condition ($t_{(136)} = 4.4$, $P < 0.001$) and the condition in which no lights were visible ($t_{(132)} = 6.4$, $P < 0.001$; see Supplementary Information). In other words, the background visual motion caused a deviation in the reaching movement, and this helped to counteract a bias towards underestimating the magnitude of ego-motion.

Although retinal motion information might be useful for controlling action in cases of whole-body rotation, could it be useful in cases where just the eye moves? Because the relative positions of the hand and target do not change during an eye movement, it seems that any influence of motion on reaching movements would be counterproductive. To test this, we conducted an experiment similar to the previous one, except that subjects moved their eyes (smooth pursuit) rather than rotating their entire bodies (see Supplementary Information). Results showed that pointing movements were more accurate when there was a stationary background (providing retinal motion as the subject moved his or her eyes) than when the background was intermittently visible or not visible. Retinal motion, produced by either eye or body movements, therefore reduces a systematic bias in reaching movements.

Reaching quickly to targets seems effortless despite the frequency and even the unintentionality of ego-motion. The experiments here suggest that part of the reason for this precision and speed might be the use of visual motion signals as a means of rapidly gauging and compensating for errors that occur when the eye or body moves. □

Methods

Stimuli were presented on a CRT monitor (NEC, Itasca, Illinois, USA) with a refresh rate of 85 Hz, controlled by a Macintosh G4 computer (Apple Computer, Cupertino, California,

USA). The stimulus consisted of a pair of horizontally striped patterns (gratings) that drifted vertically in opposite directions (Fig. 1a). Each grating subtended $5.76^\circ \times 26.76^\circ$, and each was centred at $\pm 5.76^\circ$ eccentricity. The gratings had a sinusoidal luminance modulation of 0.17 cycles per degree at 85% contrast on a dark background (0.03 cd m^{-2}) and translated at 2.7 Hz. Subjects were seated 48 cm from the monitor in a darkened room with a chin rest to stabilize the head. Two of the authors and two naive subjects participated in the first experiment (a total of six additional subjects participated in the subsequent experiments). Subjects were right-handed and had normal or corrected-to-normal visual acuity. Reaching movements were recorded, by using a single infrared emitter on the tip of the index finger, with an Optotrak 3020 (Northern Digital, Inc., Waterloo, Ontario, Canada) at 250 Hz, under the control of the same computer that generated the stimuli.

A fixation bull's-eye was provided at the centre of the screen. In each trial there was a random vertical offset in the position of the bull's-eye within a range of 1.16° . Subjects were asked to fixate on the bull's-eye at all times. In each experimental trial, the two gratings (Fig. 1a) were initially stationary for ~500 ms and then began to translate in opposite directions. The initial direction of the gratings' motion was random in each trial. After a randomly determined interval between 1,050 and 1,750 ms the gratings both abruptly reversed direction and translated for an equivalent period. At varying periods of time before or after the gratings reversed direction (ISA, from 940 ms before to 470 ms after the reversal), a target bar (84.2 cd m^{-2}) was presented (as in Fig. 1a) for ~24 ms. The target subtended $2.89^\circ \times 0.29^\circ$ and could be located in one of three randomly determined positions, either vertically centred or 1.16° above or below centre. The Optotrak began recording the position of the hand immediately after the target was presented. Subjects pointed, with their right index finger, as quickly as possible to the position of the target, hitting the CRT with their finger. The distance between the hand's resting position and the target on the monitor was ~31–42 cm. The onset of the reaching movement was calculated by using a threshold hand velocity (50 mm s^{-1} for 20 ms). The end of the reaching movement was calculated as the moment that the hand's deceleration peaked (the subject hit the immobile CRT at a high velocity, and the resulting instantaneous deceleration of the hand marked the moment that the finger hit the screen). In the first experiment there were 90 trials for each of the seven possible ISAs. To measure average hand trajectories (in Figs 2–4) a minimum of 300 additional trials were collected in separate experimental sessions for each of the conditions of interest (the additional trials explain why the endpoints shown in Fig. 1b are not identical to the average endpoints shown in Figs 2–4). The average trajectory was calculated as the normalized position of the hand, every 4 ms for each of the (three) target positions, as a function of motion direction.

In the last experiment, subjects were seated in a vestibular chair that rotated sinusoidally (through $\sim 90^\circ$) at a peak frequency of $\sim 0.2 \text{ Hz}$. A chin rest was provided to immobilize the head. The head was located 20 cm from the axis of rotation. Subjects fixated on a red LED (16 cm distance) that was attached to the head (so the eye was fixed with respect to the body). Vision was monocular and the room was completely dark; subjects could not see their hand under any of the conditions. While subjects were stationary, a ~100-ms target LED (~40 cm distance) was presented in one of five randomly determined locations. Subjects began to rotate ~1 s later, clockwise and anti-clockwise on alternate trials. At a randomly determined time (1.4–2.8 s after the start of rotation), subjects were instructed to point as quickly as possible to the remembered location of the target. Endpoints of the reaching movements were determined as described above. In one condition there were 25 additional LEDs in the background that were continuously visible throughout the experiment (see Supplementary Information). In a second condition, the background LEDs were strobed at ~4 Hz (visible for ~30 ms, off for ~200 ms). In a third condition, there was no visible background. There were 100 trials in each of the three conditions. The mean endpoint position of the hand was calculated for each of the five target positions and two directions of rotation (reach error = mean endpoint minus the actual location of the target).

Received 27 February; accepted 14 April 2003; doi:10.1038/nature01693.

1. Jeannerod, M. *The Neural and Behavioural Organization of Goal-Directed Movements* (Clarendon Press, Oxford, 1988).
2. De Valois, R. L. & De Valois, K. K. Vernier acuity with stationary moving Gabors. *Vision Res.* **31**, 1619–1626 (1991).
3. Kowler, E. *Eye Movements and their Role in Visual and Cognitive Processes* (Elsevier, Amsterdam, 1990).
4. Nishida, S. & Johnston, A. Influence of motion signals on the perceived position of spatial pattern. *Nature* **397**, 610–612 (1999).
5. Ramachandran, V. S. & Anstis, S. M. Illusory displacement of equiluminous kinetic edges. *Perception* **19**, 611–616 (1990).
6. Smeets, J. B. J. & Brenner, E. Perception and action are based on the same visual information: distinction between position and velocity. *J. Exp. Psychol. Hum. Percept. Perform.* **21**, 19–31 (1995).
7. Whitney, D. & Cavanagh, P. Motion distorts visual space: shifting the perceived position of remote stationary objects. *Nature Neurosci.* **3**, 954–959 (2000).
8. Bridgeman, B., Lewis, S., Heit, G. & Nagle, M. Relation between cognitive and motor-oriented systems of visual position perception. *J. Exp. Psychol. Hum. Percept. Perform.* **5**, 692–700 (1979).
9. Wong, E. & Mack, A. Saccadic programming and perceived location. *Acta Psychol. (Amst.)* **48**, 123–131 (1981).
10. Desmurget, M. & Grafton, S. Forward modeling allows feedback control for fast reaching movements. *Trends Cogn. Sci.* **4**, 423–431 (2000).
11. Goodale, M. A., Pelisson, D. & Prablanc, C. Large adjustments in visually guided reaching do not depend on vision of the hand or perception of target displacement. *Nature* **320**, 748–750 (1986).
12. Paulignan, Y., MacKenzie, C., Marteniuk, R. & Jeannerod, M. The coupling of arm and finger movements during prehension. *Exp. Brain Res.* **79**, 431–435 (1990).
13. Pelisson, D., Prablanc, C., Goodale, M. A. & Jeannerod, M. Visual control of reaching movements without vision of the limb. II. Evidence of fast unconscious processes correcting the trajectory of the hand to the final position of a double-step stimulus. *Exp. Brain Res.* **62**, 303–311 (1986).

14. Prablanc, C. & Martin, O. Automatic control during hand reaching at undetected two-dimensional target displacements. *J. Neurophysiol.* **67**, 455–469 (1992).
15. Brenner, E. & Smeets, J. B. J. Fast responses of the human hand to changes in target position. *J. Mot. Behav.* **29**, 297–310 (1997).
16. Abrams, R. A. & Landgraf, J. Z. Differential use of distance and location information for spatial localization. *Percept. Psychophys.* **47**, 349–359 (1990).
17. Bacon, J. H., Gordon, A. & Schulman, P. H. The effect of two types of induced-motion displays on perceived location of the induced target. *Percept. Psychophys.* **32**, 353–359 (1982).
18. Bridgeman, B., Peery, S. & Anand, S. Interaction of cognitive and sensorimotor maps of visual space. *Percept. Psychophys.* **59**, 456–469 (1997).
19. Lepecq, J. C., Joun, E. & Dubon, D. The effect of linearvection on manual aiming at memorized directions of stationary targets. *Perception* **22**, 49–60 (1993).
20. Masson, G., Proteau, L. & Mestre, D. R. Effects of stationary and moving textured backgrounds on the visuo-oculo-manual tracking in humans. *Vision Res.* **35**, 837–852 (1995).
21. Sheth, B. R. & Shimojo, S. In space, the past can be recast but not the present. *Perception* **29**, 1279–1290 (2000).
22. Yamagishi, N., Anderson, S. J. & Ashida, H. Evidence for dissociation between the perceptual and visuomotor systems in humans. *Proc. R. Soc. Lond. B* **268**, 973–977 (2001).
23. Andersen, R. A., Snyder, L. H., Bradley, D. C. & Xing, J. Multimodal representation of space in the posterior parietal cortex and its use in planning movements. *Annu. Rev. Neurosci.* **20**, 303–330 (1997).
24. Buneo, C. A., Jarvis, M. R., Batista, A. P. & Andersen, R. A. Direct visuomotor transformations for reaching. *Nature* **416**, 632–636 (2002).
25. Henriques, D. Y., Klier, E. M., Smith, M. A., Lowy, D. & Crawford, J. D. Gaze-centered remapping of remembered visual space in an open-loop pointing task. *J. Neurosci.* **18**, 1583–1594 (1998).
26. Berry, M. J. 2nd, Brivanlou, I. H., Jordan, T. A. & Meister, M. Anticipation of moving stimuli by the retina. *Nature* **398**, 334–338 (1999).
27. Hayes, A. Apparent position governs contour-element binding by the visual system. *Proc. R. Soc. Lond. B* **267**, 1341–1345 (2000).
28. Snowden, R. J. Shifts in perceived position following adaptation to visual motion. *Curr. Biol.* **8**, 1343–1345 (1998).
29. Whitaker, D., McGraw, P. V. & Pearson, S. Non-veridical size perception of expanding and contracting objects. *Vision Res.* **39**, 2999–3009 (1999).
30. Blouin, J., Gauthier, G. M., van Donkelaar, P. & Vercher, J. L. Encoding the position of a flashed visual target after passive body rotations. *NeuroReport* **6**, 1165–1168 (1995).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. Ladich, D. Pulham, C. Thomas, L. Van Cleeff, E. Veinsreideris and H. Yang. This work was supported by grants from US National Institutes of Health/National Eye Institute to D.W., and from Canadian Institutes of Health Research and the Canada Research Chairs Program to M.A.G.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.W. (dvw@uwo.ca).

Abundant gene conversion between arms of palindromes in human and ape Y chromosomes

Steve Rozen*, Helen Skaletsky*, Janet D. Marszalek*, Patrick J. Minx†, Holland S. Cordum†, Robert H. Waterston†, Richard K. Wilson† & David C. Page*

* Howard Hughes Medical Institute, Whitehead Institute, and Department of Biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA

† Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Boulevard, St Louis, Missouri 63108, USA

Eight palindromes comprise one-quarter of the euchromatic DNA of the male-specific region of the human Y chromosome, the MSY¹. They contain many testis-specific genes and typically exhibit 99.97% intra-palindromic (arm-to-arm) sequence identity¹. This high degree of identity could be interpreted as evidence that the palindromes arose through duplication events that occurred about 100,000 years ago. Using comparative sequencing in great apes, we demonstrate here that at least six of these MSY

palindromes predate the divergence of the human and chimpanzee lineages, which occurred about 5 million years ago. The arms of these palindromes must have subsequently engaged in gene conversion, driving the paired arms to evolve in concert. Indeed, analysis of MSY palindrome sequence variation in existing human populations provides evidence of recurrent arm-to-arm gene conversion in our species. We conclude that during recent evolution, an average of approximately 600 nucleotides per newborn male have undergone Y–Y gene conversion, which has had an important role in the evolution of multi-copy testis gene families in the MSY.

The human MSY palindromes, designated P1–P8, are surprisingly large, with arm lengths that range from 9 kilobases (kb; P7) to 1.45 megabases (Mb; P1) (see Table 2 and Figs 2, 3 and 5 of the accompanying manuscript¹). The paired arms of each palindrome are separated by a non-duplicated spacer that measures 2–170 kb in length. Fifteen gene and transcript families have been identified in the palindrome arms (none in the spacers), and all seem to be expressed predominantly or exclusively in testes¹. Similar to the palindrome arms in which they reside, these gene families are characterized by extremely low sequence divergence between the copies found in a single Y chromosome.

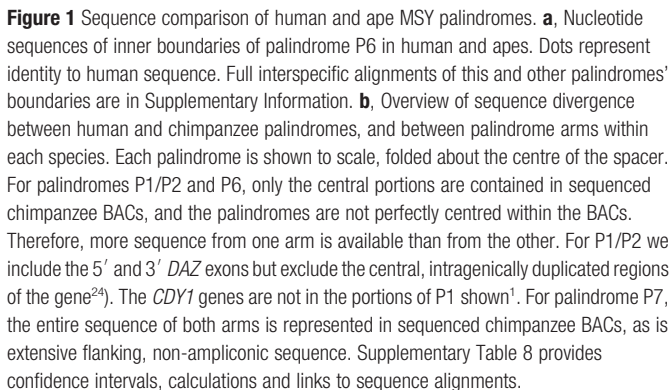
The DAZ gene family of the MSY resides exclusively in the arms of palindromes P1 and P2 (ref. 2). Near identity between DAZ copies in a single Y chromosome led some investigators to conclude, based on molecular clock reasoning, that DAZ gene amplification had occurred only within the last 200,000 years³. However, multiple Y-linked copies of DAZ also exist in apes and Old World monkeys^{3–6}. This suggests that palindromes P1 and P2, which contain the DAZ genes, might predate the divergence of humans from other primate lineages. This may be true for the other MSY palindromes as well. In that case, the near identity observed between palindrome arms could be the consequence of gene conversion—“the non-reciprocal transfer of information from one DNA duplex to another”. Gene conversion sometimes involves transfer between repeated sequences on the same chromosome⁸.

To test the ancient origins/gene-conversion hypothesis, we looked for evidence that MSY palindromes were present in the common ancestor of humans and chimpanzees. Specifically, we searched for orthologues of the eight human palindromes in chimpanzees (*Pan troglodytes*), bonobos (pygmy chimpanzee, *Pan paniscus*) and gorillas (*Gorilla gorilla*). In each species, and for each palindrome, we attempted to amplify, by polymerase chain reaction (PCR), and sequence the two inner boundaries (between spacer and arms) and the two outer boundaries (between arms and surrounding sequences). We successfully amplified both inner boundaries in multiple palindromes (Table 1). In all of these cases, the PCR products were observed only when male genomic DNAs were used as templates, and never when using female genomic DNAs (data not shown). This implies that the PCR products were amplified from the male-specific regions of the great ape Y chromosomes. In all cases, the boundary sequences were essentially identical in humans and great apes (Fig. 1a; see also Supplementary Information). Only for P7 did we successfully amplify both outer boundaries (in chimpanzee and bonobo). These findings suggested that: (1) most palindromes found in the modern human MSY were

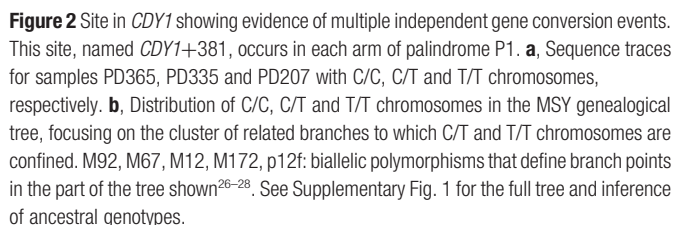
Table 1 Ape MSY palindromes confirmed by sequencing of inner boundaries

Species	Palindrome					
	P1	P2	P4	P6	P7	P8
Chimpanzee	Y	Y	–	Y	Y	Y
Bonobo	Y	Y	–	Y	Y	–
Gorilla	–	–	Y	Y	–	–

Sequence alignments provided in Supplementary Information. Y indicates confirmation of MSY palindromes.



To enable detailed comparisons of human and chimpanzee palindromes, we screened a male chimpanzee genomic bacterial artificial chromosome (BAC) library for clones homologous to the inner boundaries of human palindromes P1–P8. We identified and then sequenced chimpanzee BACs corresponding to palindromes P1, P2, P6 and P7. (The BAC library provided only one- to twofold coverage, on average, of chimpanzee MSY sequences and thus was not expected to contain all boundaries of MSY palindromes.) Comparative sequence analysis confirmed the structural similarity of the human and chimpanzee palindromes and, by inference, their common ancestry (Fig. 1b; see Supplementary Information for complete sequence alignments). We observed 1.44% sequence divergence, on average, between orthologous palindrome arms in human and chimpanzee (Fig. 1b and Table 2). Such divergence between species probably reflects the simple accumulation of neutral mutations in the human and chimpanzee lineages after



If gene conversion between palindrome arms was responsible for our findings, it might leave traces in the recent genealogy of the human MSY. In particular, we might find evidence that single nucleotide differences between the two arms of a human MSY palindrome had been eliminated by gene conversion. Examination of two *CDY* genes—one in each arm of palindrome P1—revealed a duplicated site of sequence variation that fulfilled this prediction. By sequencing this duplicated site in diverse, unrelated men, we identified some Y chromosomes with a C at this site in both arms of P1 (C/C chromosomes), other chromosomes with a C in one arm and a T in the second arm (C/T chromosomes), and other chromosomes with a T in both arms (T/T chromosomes; Fig. 2a). We confirmed these findings using a PCR/restriction-digestion assay (Supplementary Fig. 4). This single nucleotide substitution occurs at nucleotide 381 of the *CDY* coding region but does not alter the predicted amino acid sequence.

NATURE | VOL 423 | 19 JUNE 2003 | www.nature.com/nature

Table 2 Sequence divergence in MSY palindromes

Comparison	Length (kb)	Divergence			
		(%)	95% CI	<i>P</i> -value versus palindrome spacer	<i>P</i> -value versus non-amplificonic MSY*
Human versus chimpanzee					
Palindromic arm†	123	1.44	1.37–1.51	2.2×10^{-16}	7.2×10^{-16}
Palindromic spacer	61	2.26	2.14–2.38	–	2.4×10^{-14}
Non-amplificonic MSY*	287	1.79	1.75–1.84	–	–
Human (arm-to-arm)	82	0.021	0.012–0.035	–	–
Chimpanzee (arm-to-arm)‡	76	0.028	0.017–0.042	–	–

See Supplementary Tables 6–8 for details.

*Includes non-amplificonic sequence flanking P7 and sequence of an additional, non-amplificonic chimpanzee BAC.

†Some of the chimpanzee sequences are represented in a single copy in a sequenced chimpanzee BAC; results are unchanged if analysis is restricted to the portions of arms for which two copies were sequenced in chimpanzee.

‡The chimpanzee arm sequences compared are shorter than the corresponding human sequences because of deletions in the chimpanzee arms relative to human.

some) by a C → T substitution in one arm of palindrome P1.

In three of this cluster's five branches, we observed T/T as well as C/T chromosomes (Fig. 2; see also Supplementary Fig. 1). This finding is readily explained by gene conversion in a C/T chromosome—the ancestral chromosome for this cluster—replacing the C in one arm of palindrome P1 with the T in the other arm. The data reveal at least three such gene-conversion events—one in each of the branches that have T/T chromosomes (Fig. 2b). In one of these branches, we also observed C/C chromosomes alongside C/T and T/T chromosomes (Fig. 2b). Here we surmise that gene conversion in a C/T chromosome replaced the T in one arm of P1 with the C in the other arm. Thus, during recent human history, gene conversion in C/T chromosomes has used either the C copy or the T copy as template. In addition, we investigated two other duplicated sites of sequence variation, and at both sites we found evidence of recurrent gene conversion during recent human history (Supplementary Figs 2 and 3).

How frequently does gene conversion occur in the MSY palindromes? Near uniformity of arm-to-arm sequence divergence in both human and chimpanzee palindromes (Table 2 in ref. 1 and Fig. 1b) suggests a steady-state balance between new mutations that create differences between arms, and gene-conversion events that erase these differences. Accordingly, we can calculate the rate of gene conversion needed to maintain the observed divergence in the face of new mutations. Let μ be the human MSY mutation rate, 1.6×10^{-9} substitutions per nucleotide per year (see Methods). Let d be the observed divergence between human MSY palindrome arms (3×10^{-4} substitutions per duplicated nucleotide), and let c be the (unknown) rate of gene conversion (in both directions combined) per duplicated nucleotide per year. Differences between arms are introduced at a rate of 2μ (as a mutation in either arm creates a difference between arms), and homogenized at a rate of cd . Thus, at steady state, $cd = 2\mu$. Then $c = 2\mu/d = 2 \times 1.6 \times 10^{-9} / 3 \times 10^{-4} = 1.1 \times 10^{-5}$ gene conversions per duplicated nucleotide per year. For a 20-year human generation, this corresponds to a rate of 2.2×10^{-4} conversions per duplicated nucleotide per generation, comparable to rates estimated directly in a mouse transgenic system¹⁰. Over the 5.4 Mb in human MSY palindromes (2.7×10^6 duplicated nucleotides), then, an average of about 600 duplicated nucleotides have undergone arm-to-arm gene conversion for every son born in recent human evolution. Most of these conversions would have involved two identical DNA sequences, and thus their products would be unobservable. The inferred kinetics of gene conversion in MSY palindromes is especially striking because the MSY was previously viewed as recombinationally inert under normal circumstances: it was known previously as the non-recombining region, or NRY.

At present, we do not know whether gene conversion in MSY palindromes occurs during meiosis, mitosis, or both. It may involve homology-directed double-strand break repair, as in gene conversion between homologous chromosomes or sister chromatids¹¹. An

interesting observation is that human–chimpanzee divergence is significantly reduced in MSY palindrome arms as compared with other MSY sequences examined (Table 2). This reduction is evident even when comparing Alu and other interspersed repeat sequences that are presumed to be of little functional consequence (Supplementary Table 1). Thus, the reduced rate of evolution in palindrome arms does not seem to be due to selective constraints. A weak directional bias in gene conversion, favouring restoration of the original sequence, might account for these observations.

Our finding of abundant gene conversion in MSY palindromes raises questions about the molecular-clock dating of other segmental duplications in the human genome¹². Some of these were interpreted as being of recent origin based on low copy-to-copy divergence¹³. In other cases, however, analysis by Southern blots^{14,15} or quantitative PCR¹⁶ indicated that these duplications exist in great apes as well as in humans. Thus, these duplications might well represent conserved genomic organizations subject to gene conversion and concerted evolution. In the case of human X-chromosomal colour vision genes, 2 kb of comparative sequence data confirm concerted evolution^{17,18}. Our current findings, taken together with these previous results, raise the possibility that gene conversion in primate genomes could be much more pervasive than previously thought.

Finally, we note a strong association between gene conversion and MSY testis genes. In humans, all genes in MSY palindromes seem to be expressed predominantly or exclusively in testes, and most MSY genes with this expression pattern occur in palindromes¹. Given the abundance of gene conversion in palindromes, we infer that Y–Y gene conversion has accompanied and shaped the evolution of multi-copy testis gene families in the MSY. Perhaps some selective advantage stemmed from the palindromic duplication of MSY testis genes during human evolution. If so, has Y–Y gene conversion had a role in that advantage? Has it allowed genes in palindromes to resist, or at least retard, the evolutionary decay that is a hallmark of Y chromosome evolution¹⁹? This could explain the observation, as reported in the accompanying paper, that intact testis-specific genes tend to be located in palindrome arms whereas non-functional copies of these genes seem to be distributed randomly (see Table 4 in ref. 1). A full understanding of the functional and evolutionary significance of our findings will require further study in primates and other mammals. □

Methods

Estimating the MSY mutation rate

We estimated the MSY mutation rate in the human lineage based on the data and analysis in ref. 20, and an estimate of 5.5 million years ago for the most recent common ancestor of humans and chimpanzees²¹. The result is 1.6×10^{-9} substitutions per nucleotide per year (Supplementary Fig. 5).

PCR amplification and sequencing of palindrome boundaries

Supplementary Table 2 lists the PCR primers and conditions used to amplify palindrome boundaries. Supplementary Table 3 provides GenBank accession numbers for the chimpanzee, bonobo and gorilla sequences obtained.

Identification and sequencing of chimpanzee BACs

We screened high-density filters from the RPCI-43 male chimpanzee BAC library²² (BACPAC resources) using hybridization probes designed to detect sequences (1) near the inner boundaries of palindromes P1–P6 and P8; (2) near P7; and (3) from a non-ampliconic region of the human MSY. STS content and BAC-end sequences confirmed that, among the candidate BACs identified by hybridization, six contained the central portions of orthologues to human MSY palindromes. The BACs were sequenced as previously described². Supplementary Table 4 provides descriptions of the sequenced BACs and their GenBank accession numbers.

Sequence analysis

Sequences were aligned with CLUSTAL W using default parameters²³. In a few cases, the resulting alignments were adjusted manually. All alignments are provided as Supplementary Information.

Typing nucleotide variants in palindrome arms

The sites studied were *CDY1*+381 (Fig. 2 and Supplementary Fig. 1), *CDY1*–84 (Supplementary Fig. 2), and sY586 (Supplementary Fig. 3). sY586 was genotyped as previously described²⁴. PCR primers and conditions for amplifying *CDY1*+381 (sY1313) and *CDY1*–84 (sY1314) have been deposited in GenBank (accession numbers G73596 and G73597, respectively). When typing *CDY1*+381 by sequencing, 'primer A' in GenBank G73596 served as the sequencing primer. *CDY1*–84 was typed by sequencing using 'primer B' in GenBank G73597.

For the samples that showed evidence of gene conversion (Fig. 2 and Supplementary Figs 1–3), we excluded the possibility of deletion of one copy of the variant site as discussed in Supplementary Note 1.

Steady-state balance between mutations and gene-conversion

To show that the combined action of mutation and gene conversion results in a steady-state level of arm-to-arm divergence, we use the following recursion: $d_{n+1} = (1 - c_g)d_n + 2\mu_g$ where d_n is the sequence divergence between repeat copies at generation n , μ_g is the mutation rate per nucleotide per generation, and c_g is the gene conversion rate per duplicated nucleotide per generation. We presume that $d_0 = 0$, corresponding to no differences between sequence copies immediately after the initial duplication event. However, as $1 - c_g < 1$, $\lim_{n \rightarrow \infty} d_n = 2\mu_g/c_g$ for any value of d_0 small enough to support c_g . Because μ_g and d are very small, mutations almost never occur at sites that already differ between the two palindrome arms, and this possibility can be ignored. As shown in Supplementary Note 2, our analysis is a special case of Ohta's analysis²⁵.

Received 10 March; accepted 7 April 2003; doi:10.1038/nature01723.

- Skaletsky, H. *et al.* The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**, 825–837 (2003).
- Kuroda-Kawaguchi, T. *et al.* The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. *Nature Genet.* **29**, 279–286 (2001).
- Agulnik, A. I. *et al.* Evolution of the DAZ gene family suggests that Y-linked DAZ plays little, or a limited, role in spermatogenesis but underlines a recent African origin for human populations. *Hum. Mol. Genet.* **7**, 1371–1377 (1998).
- Reijo, R. *et al.* Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene. *Nature Genet.* **10**, 383–393 (1995).
- Glaser, B. *et al.* Simian Y chromosomes: species-specific rearrangements of DAZ, RBM, and TSPY versus contiguity of PAR and SRY. *Mamm. Genome* **9**, 226–231 (1998).
- Makova, K. D. & Li, W. H. Strong male-driven evolution of DNA sequences in humans and apes. *Nature* **416**, 624–626 (2002).
- Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J. & Stahl, F. W. The double-strand-break repair model for recombination. *Cell* **33**, 25–35 (1983).
- Jackson, J. A. & Fink, G. R. Gene conversion between duplicated genetic elements in yeast. *Nature* **292**, 306–311 (1981).
- Underhill, P. A. *et al.* The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann. Hum. Genet.* **65**, 43–62 (2001).
- Murti, J. R., Bumbulis, M. & Schimenti, J. C. High-frequency germ line gene conversion in transgenic mice. *Mol. Cell. Biol.* **12**, 2545–2552 (1992).
- Johnson, R. D. & Jasin, M. Sister chromatid gene conversion is a prominent double-strand break repair pathway in mammalian cells. *EMBO J.* **19**, 3398–3407 (2000).
- Bailey, J. A. *et al.* Recent segmental duplications in the human genome. *Science* **297**, 1003–1007 (2002).
- International Human Genome Sequencing Consortium Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
- Small, K., Iber, J. & Warren, S. T. Emerin deletion reveals a common X-chromosome inversion mediated by inverted repeats. *Nature Genet.* **16**, 95–99 (1997).
- Aradhy, S. *et al.* Multiple pathogenic and benign genomic rearrangements occur at a 35 kb duplication involving the *NEMO* and *LAG2* genes. *Hum. Mol. Genet.* **10**, 2557–2567 (2001).
- Rochette, C. F., Gilbert, N. & Simard, L. R. *SMN2* gene duplication and the emergence of the *SMN2* gene occurred in distinct hominids: *SMN2* is unique to *Homo sapiens*. *Hum. Genet.* **108**, 255–266 (2001).
- Deeb, S. S., Jorgensen, A. L., Battisti, L., Iwasaki, I. & Motulsky, A. G. Sequence divergence of the red and green visual pigments in great apes and humans. *Proc. Natl Acad. Sci. USA* **91**, 7262–7266 (1994).
- Zhou, Y.-H. & Li, W.-H. Gene conversion and natural selection in the evolution of X-linked color vision genes in higher primates. *Mol. Biol. Evol.* **18**, 780–783 (1996).
- Charlesworth, B. & Charlesworth, D. The degeneration of Y chromosomes. *Phil. Trans. R. Soc. Lond. B* **355**, 1563–1572 (2000).
- Bohossian, H. B., Skaletsky, H. & Page, D. C. Unexpectedly similar rates of nucleotide substitution found in male and female hominids. *Nature* **406**, 622–625 (2000).
- Kumar, S. & Hedges, S. B. A molecular timescale for vertebrate evolution. *Nature* **392**, 917–920 (1998).

- Fujiyama, A. *et al.* Construction and analysis of a human-chimpanzee comparative clone map. *Science* **295**, 131–134 (2002).
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
- Saxena, R. *et al.* Four DAZ genes in two clusters found in AZFc region of human Y chromosome. *Genomics* **67**, 256–267 (2000).
- Ohta, T. Allelic and nonallelic homology of a supergene family. *Proc. Natl Acad. Sci. USA* **79**, 3251–3254 (1982).
- Casanova, M. *et al.* A human Y-linked DNA polymorphism and its potential for estimating genetic and evolutionary distance. *Science* **230**, 1403–1406 (1985).
- Underhill, P. A. *et al.* Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res.* **7**, 996–1005 (1997).
- Shen, P. *et al.* Population genetic implications from sequence variation in four Y chromosome genes. *Proc. Natl Acad. Sci. USA* **97**, 7354–7359 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank R. K. Alagappan and L. G. Brown for technical contributions; N. A. Ellis, M. F. Hammer, T. Jenkins and P. A. Underhill for assistance with genealogical studies; H. M. McClure and Yerkes Regional Primate Research Center for samples; C. Disteche, A. E. Donnenfeld, J. H. Hersh, T. Jenkins, P. G. McDonough, B. McGillivray, R. D. Oates, P. Patrizio, R. Rosenfield, L. Shapiro, S. Silber, M. C. Summers, J. Weissenbach, B. Whitmire and S. Yang for patient samples; and J. E. Alfoldi, B. Charlesworth, A. G. Clark, J. Koubova, J. Lange, B. Levy, T. L. Orr-Weaver, S. Repping, W. R. Rice and J. Saionz for comments on the manuscript. This work was supported by the National Institutes of Health and the Howard Hughes Medical Institute.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.C.P. (page_admin@wi.mit.edu). All new DNA sequences and STSs were submitted to GenBank with accession numbers AC139189–AC139194 (chimpanzee BACs), AY090860–AY090881 (palindrome boundary sequences in apes), and G73582–G73595 (STS for amplifying palindrome boundaries); see Supplementary Information for details.

Fibronectin requirement in branching morphogenesis

Takayoshi Sakai, Melinda Larsen & Kenneth M. Yamada

Craniofacial Developmental Biology and Regeneration Branch, National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, Maryland 20892-4370, USA

Many organs, including salivary glands, lung and kidney, are formed during embryonic development by epithelial branching. In branching morphogenesis, repetitive epithelial cleft and bud formation create the complex three-dimensional branching structures characteristic of many organs^{1–3}. Although the mechanisms are poorly understood, one might involve the site-specific accumulation of some regulatory protein. Here we show that the extracellular matrix protein fibronectin^{4,5} is essential for cleft formation during the initiation of epithelial branching. Fibronectin messenger RNA and fibrils appeared transiently and focally in forming cleft regions of submandibular salivary-gland epithelia, accompanied by an adjacent loss of cadherin localization. Decreasing the fibronectin concentration by using small interfering RNA and inhibition by anti-fibronectin or anti-integrin antibodies blocked cleft formation and branching. Exogenous fibronectin accelerated cleft formation and branching. Similar effects of fibronectin suppression and augmentation were observed in developing lung and kidney. Mechanistic studies revealed that fibrillar fibronectin can induce cell–matrix adhesions on cultured human salivary epithelial cells with a local loss of cadherins at cell–cell junctions. Thus, fibronectin expression is required for cleft formation in branching

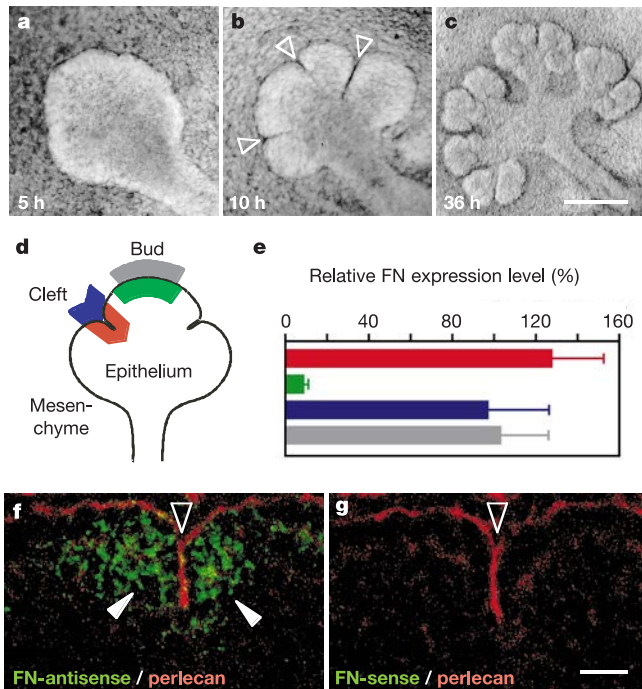


Figure 1 Expression of fibronectin mRNA in branching morphogenesis of mouse embryonic salivary gland. **a–c**, Submandibular glands from E12.5 mice were cultured on a membrane for 5 h (**a**), 10 h (**b**) or 36 h (**c**). Glands cultured for 10 h show buds separated by narrow, deep clefts (open triangles). Scale bar, 100 μ m. **d**, Diagram of salivary epithelial regions isolated by laser microdissection: cleft epithelium (red) and bud epithelium (green); mesenchyme adjacent to a cleft (blue) or bud (grey). **e**, Quantitative real-time RT–PCR analysis of expression of fibronectin mRNA at each site. Bars indicate s.e.m. **f, g**, Expression of epithelial fibronectin mRNA as determined by whole-mount fluorescence *in situ* hybridization imaged by confocal microscopy. Open triangles indicate clefts; filled triangles point to sites of fibronectin expression (green). Perlecan protein immunostaining was used as a marker of basement membrane (red). Fibronectin transcripts were identified by the fibronectin antisense probe (**f**), and the absence of fibronectin sense probe hybridization demonstrated specificity (**g**). Scale bar, 20 μ m.

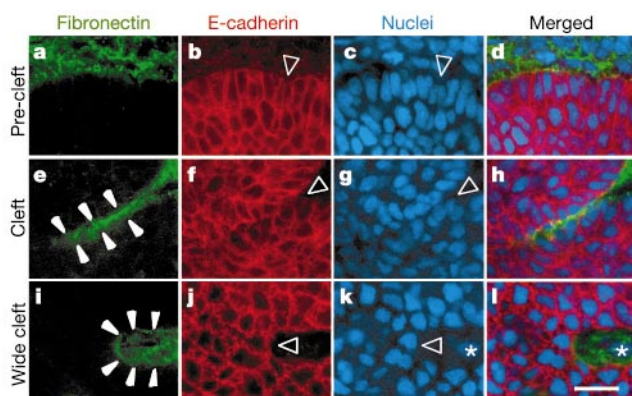


Figure 2 Expression of fibronectin protein during branching morphogenesis, shown by staining with anti-fibronectin and imaging by confocal microscopy. Open triangles indicate clefts, and filled triangles outline regions of fibronectin expression in these single confocal sections. Fibronectin staining is prominent in mesenchyme but minimal in epithelium before cleft formation (**a**), accumulates during cleft formation (**e**), and gradually decreases in widening clefts (**i**). Triple-staining of the same confocal section with anti-E-cadherin (red; **b, f, j**) and SYBR green I (blue; **c, g, k**) is shown merged in panels **d, h** and **l**. A mesenchymal cell (asterisk) has appeared in the widening cleft (**k, l**). Scale bar, 20 μ m.

morphogenesis associated with the conversion of cell–cell adhesions to cell–matrix adhesions.

Although direct mechanisms remain to be described, the overall process of branching morphogenesis depends on epithelial–mesenchymal interactions and growth factors^{1–3,6}. The mouse submandibular salivary gland has been a classical model for analysing branching morphogenesis: extracellular matrix, growth factors and actin microfilaments, but not apoptosis, have key functions in the formation of salivary-gland clefts and buds^{7–10}. Isolated epithelium without mesenchyme can still undergo branching¹¹. Although epithelial basement membrane components such as laminin and proteoglycans are required for salivary branching and epithelial functions in general^{12–14}, it is important to identify locally synthesized regulatory proteins in the epithelium that directly regulate the sites and steps of branching. Type III collagen accumulates at clefts¹⁵, and the formation of type III collagen fibrils can be controlled by fibronectin¹⁶. This protein promotes cell adhesion, migration and signalling^{4,5,17} and is widely expressed in developing tissues, including branching organs^{5,8}. We propose that a novel process of local, developmentally programmed expression of epithelial cell fibronectin might regulate branching morphogenesis, for example by converting cell–cell adhesions to cell–matrix adhesions.

Branching morphogenesis of developing salivary glands is initiated by the formation of shallow clefts in a globular bud, which deepen to generate new buds (Fig. 1a, b). The basal portions of the buds then elongate to form cylindrical stalks, and these steps are repeated successively to generate a progressively complex, branched structure^{1,2,8} (Fig. 1c). We first searched for differential expression patterns of fibronectin mRNA. To quantify local fibronectin mRNA amounts, we used laser microdissection of cryostat sections of developing submandibular glands to isolate specific cell

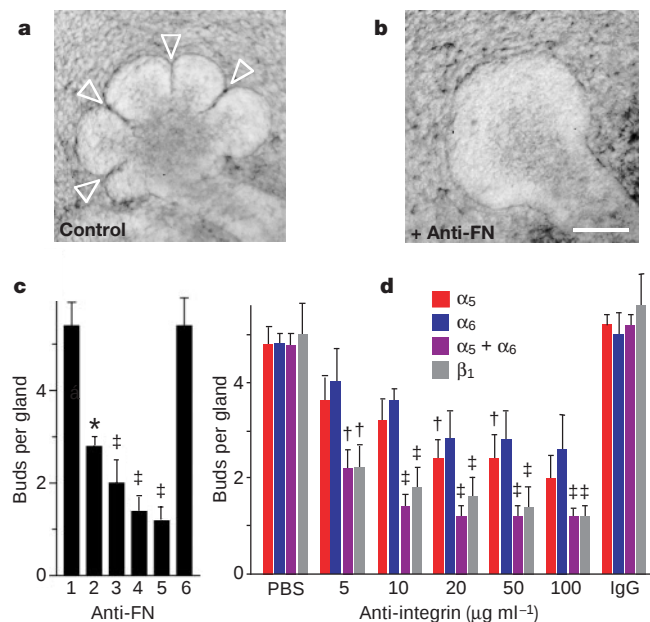


Figure 3 Inhibition of branching by anti-fibronectin antibody. **a, b**, Phase-contrast images of living E12.5 salivary glands cultured for 10 h after treatment with PBS control (**a**) or anti-fibronectin R745 antibody diluted 1:50 (**b**). Open triangles indicate clefts. Scale bar, 100 μ m. **c**, Antibody inhibition by R745 was quantified by counting numbers of buds per gland. Bar 1, PBS-treated glands (control); bars 2–5, R745-treated glands at dilutions of 1:500 (bar 2), 1:200 (bar 3), 1:100 (bar 4) and 1:50 (bar 5), and control rabbit serum diluted 1:50 (bar 6). **d**, Inhibition of branching by monoclonal antibodies against integrins compared with corresponding control IgGs. Bars indicate s.e.m. Data were analysed by Bonferroni's multiple comparison test. Asterisks, $P < 0.01$; daggers, $P < 0.05$; double daggers, $P < 0.001$ compared with controls.

populations from the clefts and buds of the salivary epithelium, and also mesenchyme adjacent to these sites (Fig. 1d). Although fibronectin is known to be produced in large amounts by mesenchymal cells, we found unexpectedly high concentrations of fibronectin mRNA in cleft epithelial cells: analysis by quantitative real-time polymerase chain reaction with reverse transcription (real-time RT-PCR) showed that cleft epithelial cells expressed 16-fold more fibronectin mRNA than bud epithelium, reaching concentrations as high as or higher than in mesenchyme (Fig. 1e). In contrast, both cleft and bud epithelia expressed the same amounts of mRNA for ribosomal protein L19 (see Supplementary Information 1).

We used *in situ* hybridization to characterize the spatial pattern of expression of fibronectin mRNA in early clefts. The epithelial cells adjacent to forming clefts expressed fibronectin mRNA, whereas all other epithelial regions showed little or no such expression (Fig. 1f, g). Immunofluorescence analysis for fibronectin with methods optimized for staining both epithelium and mesenchyme revealed minimal amounts of this protein in salivary epithelium before cleft formation, whereas concentrations were high in the mesenchyme, as described previously¹⁸ (Fig. 2a–d). Although little fibronectin was observed near the earliest epithelial dimples formed during cleft initiation, intense staining for fibrils of fibronectin was routinely observed in narrow clefts of branching salivary epithelium (Fig. 2e–h). No cells (such as mesenchyme) were detected by nuclear staining in these fibronectin-rich, narrow clefts (Fig. 2g), which were $2.7 \pm 0.4 \mu\text{m}$ wide (mean $\pm$ s.e.m.).

We also discovered an associated decrease in amounts of the cell–cell adhesion protein E-cadherin¹⁹ in the salivary epithelial cells immediately adjacent to the fibronectin fibrils (Fig. 2f). Moreover,

expression of E-cadherin mRNA in laser-microdissected cleft epithelium was sixfold lower than in bud epithelium as determined by real-time RT-PCR (see Supplementary Information 1). As clefts broadened, an apparent decrease in intensity of staining for fibronectin was observed (Fig. 2i–l); occasional mesenchymal cells could be observed entering fibronectin-containing cleft regions more than $10\text{--}15 \mu\text{m}$ wide (Fig. 2k). A similar localization of fibronectin in clefts of mesenchyme-free epithelial cultures was observed, indicating that fibronectin synthesized mesenchymally is not necessary for fibronectin localization in forming clefts (see Supplementary Information 2d–o).

We hypothesized that the transient local expression of epithelial fibronectin mRNA in presumptive cleft regions resulted in the accumulation of a fibronectin protein matrix that promoted cleft formation during branching morphogenesis. As the first direct functional test, we inhibited fibronectin with polyclonal or monoclonal antibodies. In organ cultures, anti-fibronectin inhibited salivary cleft formation and branching in a dose-dependent manner (Fig. 3a–c). Although laminin and the laminin receptor $\alpha_6\beta_1$ integrin are known to have functions in salivary branching morphogenesis¹³, our findings suggested an additional function for the fibronectin receptor $\alpha_5\beta_1$. Antibodies against the integrin α_5 subunit, and also against the β_1 and α_6 subunits, produced substantial dose-dependent inhibition of salivary branching (Fig. 3d). Combining antibodies against α_5 and α_6 inhibited branching more effectively than each antibody separately. These observations indicate that fibronectin and its major $\alpha_5\beta_1$ integrin receptor might be required for cleft formation and branching, together with the known roles for basement membrane laminin and its receptor $\alpha_6\beta_1$.

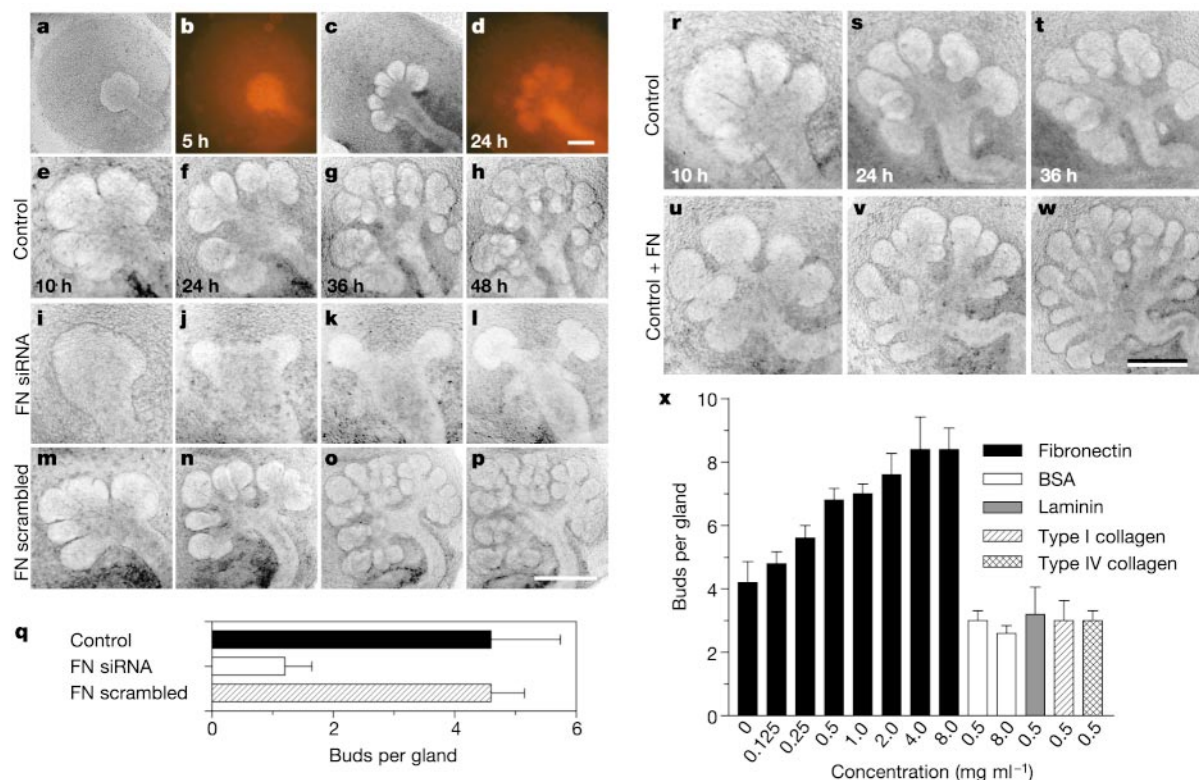


Figure 4 Effects of fibronectin siRNA and exogenous fibronectin on salivary branching morphogenesis. **a–q**, Inhibition of cleft formation by fibronectin siRNA. Cy3-labelled fibronectin siRNA (**a, b**) or a Cy3-labelled generic control siRNA (Dharmacon Scramble I; **c, d**) were transfected into salivary-gland cultures with the use of Oligofectamine, and glands were cultured on a membrane for 5 h (**a, b**) or 24 h (**c, d**). Phase-contrast images show living glands (**a, c**), and fluorescent microscopy images show Cy3-labelled siRNAs (**b, d**). Controls and fibronectin siRNA inhibition of cleft formation at 10 h (**e, i, m**), 24 h

(**f, j, n**), 36 h (**g, k, o**) and 48 h (**h, l, p**): control (PBS; **e–h**), fibronectin siRNA (**i–l**) and scrambled fibronectin siRNA (**m–p**). Scale bars, 100 μm . **q**, Quantification by counting numbers of buds per gland at 10 h. **r–x**, Exogenous fibronectin stimulates branching. Phase-contrast images show salivary glands at 10 h (**r, u**), 24 h (**s, v**) and 36 h (**t, w**) cultured without (**r–t**) or with (**u–w**) added plasma fibronectin (2 mg ml⁻¹). Scale bar, 100 μm . **x**, Fibronectin added for 10 h to otherwise untreated glands induces branching in a dose-dependent manner, but other matrix proteins and BSA do not.

To test whether the observed focal expression of fibronectin mRNA was required for branching morphogenesis, we generated a fibronectin knock-down phenotype in salivary-gland organ culture with the use of small interfering RNA (siRNA)²⁰. We designed siRNA oligonucleotides for the mouse fibronectin gene by using controls with the same nucleotides arranged in a scrambled order and a generic commercial siRNA control. Transfection was performed with Oligofectamine with a final concentration of 500 nM siRNA duplex in culture medium. Initial tests of uptake revealed that Cy3-labelled siRNA (fibronectin and a control duplex) was not only readily incorporated into glands but also showed preferential accumulation in epithelia (Fig. 4a–d). This unexpected finding indicates that siRNA approaches might be particularly effective for testing epithelial functions in intact embryonic glands *in vitro*,

which has been technically difficult in the salivary-gland system when using adenoviral and electroporation approaches (M.L., unpublished data).

Treatment with fibronectin siRNA resulted in the abolition or a substantial decrease in cleft formation in all transfected salivary glands at 10–48 h (Fig. 4i–l). The non-branched buds seemed enlarged compared with untreated controls, indicating that inhibition of branching occurred without effects on epithelial volume. The treatment was further confirmed to be non-toxic by reversing the inhibition in washout experiments (see Supplementary Information 3a–f). Equal concentrations of control siRNA consisting of the same fibronectin sequence that had been scrambled, or an irrelevant scrambled sequence, had no effects on salivary morphological development (Fig. 4m–p, and data not shown), with the same number and morphology of buds as untreated controls (Fig. 4e–h, q). Evidence that the fibronectin siRNA treatment had effectively targeted fibronectin was provided by immunofluorescence (see Supplementary Information 3g–r). The fibronectin siRNA treatment substantially decreased the localization of fibronectin shown by antibodies in both cleft epithelium and mesenchyme compared with controls. Intermediate effects on branching were observed by using less siRNA (T.S., unpublished observation).

Exogenous fibronectin was added to organ cultures to replace the fibronectin depleted by siRNA. Experimental supplementation of fibronectin successfully rescued branching morphogenesis in siRNA-treated glands (see Supplementary Information 3s–x). Fibronectin treatment stimulated branching in controls without siRNA, indicating that exogenous fibronectin might actually induce or accelerate branching morphogenesis (Fig. 4r–w). Dose-response analysis revealed progressive stimulation of branching, with increasing effects up to 4 mg ml⁻¹ (Fig. 4x). This experiment was repeated in mesenchyme-free cultures, and fibronectin also stimulated cleft formation in the absence of mesenchyme (see Supplementary Information 2p–u). Equal concentrations of bovine serum albumin (BSA) did not promote branching; laminin, collagen I, collagen III and collagen IV as matrix-protein controls also had no effect on branching morphogenesis (Fig. 4x, and Supplementary Information 4a). Fibronectin treatment accelerated cleft formation, with a doubling of the depth of clefts and an approximately twofold enhancement in the number of buds (Fig. 4u–x, and T.S., unpublished time-lapse microscopy data). Comparisons of the effects of fibronectin treatment on salivary gland cultures with two other molecules previously reported to stimulate salivary branching, the general matrix metalloproteinase inhibitor TIMP-1 (tissue inhibitor of metalloproteinases-1; ref. 21) and epidermal growth factor^{2,22}, showed that the effects of fibronectin were substantially greater (see Supplementary Information 4b, and T.S., unpublished data).

Branching morphogenesis of other organs, such as lung and especially kidney, seems to differ in the larger relative contribution of bud outgrowth compared with cleft formation during branching^{1–3}. Nevertheless, immunofluorescence comparisons of fibronectin localization during early branching of lung and kidney also showed an accumulation of fibronectin at sites of epithelial constriction and indentation (see Supplementary Information 5), supporting previous proposals of possible roles for fibronectin in branching morphogenesis of lung and kidney^{23,24}. Direct tests for roles of fibronectin, by treatment of developing lung or kidney rudiments with anti-fibronectin antibody or siRNA, inhibited branching morphogenesis; fibronectin supplementation promoted branching of both lung and kidney (see Supplementary Information 5, and data not shown).

A direct mechanism by which fibronectin might regulate cleft formation would be to induce the loss of the cadherin cell–cell adhesion molecule that we observed in cells adjacent to fibronectin fibrils (Fig. 2f). To test this possibility directly, human salivary-gland epithelial cells²⁵ (the HSG cell line) were exposed to

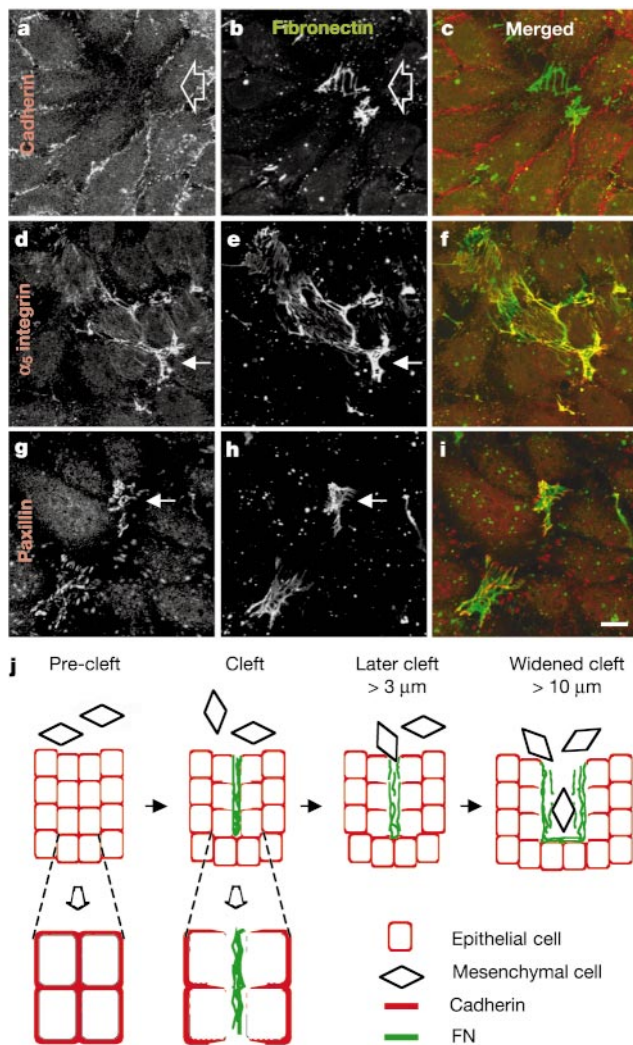


Figure 5 Local suppression of cadherin localization accompanying the formation of cell–matrix adhesions to cellular fibronectin in HSG epithelial cells. **a–i**, Exogenous cellular fibronectin was added to HSG cell cultures and incubated for 24 h. Confocal immunofluorescence images show localization of cadherins (**a**), α_5 integrin (**d**), paxillin (**g**) and fibronectin (**b**, **e**, **h**) compared with merged images (**c**, **f**, **i**). Open arrows show the area of decreased cadherins at the site of accumulation of fibronectin fibrils. Arrows show co-localization in adhesion complexes of α_5 integrin (**d**) and fibronectin (**e**), or paxillin (**g**) and fibronectin (**h**). Scale bar, 20 μ m. **j**, Summary diagram of the proposed mechanism of cleft formation in salivary epithelium. Dashed lines indicate an enlarged schematic view of four cells with suppressed cadherin localization at membrane sites that have switched from cadherin-based cell–cell adhesions to cell–matrix adhesions with fibrillar fibronectin.

pre-aggregated cellular fibronectin and examined for altered cadherin patterns. Exogenous cellular fibronectin was organized into fibrillar aggregates on the upper cell surface, and it induced local cell–matrix adhesions containing the $\alpha_5\beta_1$ integrin and cytoskeletal protein paxillin (Fig. 5d–i). Importantly, cadherin localization was suppressed at membrane sites adjacent to the fibronectin–integrin–paxillin complexes (Fig. 5a–c), indicating a local suppression of cadherin localization in cell–cell adhesions accompanying the formation of cell–matrix adhesions.

Our data show the transient, local expression of fibronectin mRNA and the focal accumulation of fibronectin fibrils in narrow forming clefts during early salivary development. Although lung and kidney branchings differ morphologically, fibronectin also accumulates at sites of epithelial indentation and constriction. Experimentally suppressing fibronectin synthesis or function by siRNA or antibodies inhibits branching morphogenesis. Moreover, supplementation of fibronectin actively promotes branching morphogenesis of salivary glands, lung and kidney. Interestingly, fibronectin accumulation was not observed at the very earliest times of salivary branching, when microfilament contraction is proposed to initiate the first dimple preceding clefting⁹; fibronectin seems to function later during the phase of cleft formation and is associated with a decreased concentration of E-cadherin and separation of epithelial cells. Although fibronectin-coated beads produced only a modest (30%) decrease in global localization of cadherin²⁶, the present study reveals a striking decrease in localization immediately adjacent to sites rich in fibronectin fibrils. This local suppression of cadherin localization by the formation of fibronectin–integrin adhesion complexes on epithelial cells provides a previously unknown mechanism for the local loss of intra-epithelial adhesion needed to form deep clefts (Fig. 5j). We suggest that the reported accumulation of type III collagen in salivary-gland clefts¹⁵ might also be due to the induction of fibronectin, because fibronectin can regulate collagen polymerization¹⁶. Collagen is thought to be essential for stabilizing clefts^{1,27}, whereas growth factors such as fibroblast growth factor promote bud extension without affecting cleft formation¹⁰. Although bud extension itself seems particularly prominent in lung and kidney morphogenesis, the present study nevertheless implicates fibronectin in developmental branching in these tissues, indicating the potential general importance of developmentally regulated, transient, focal patterns of fibronectin expression and function. □

Methods

Submandibular-gland culture

Submandibular salivary glands (SMGs) dissected from embryonic day 12.5 (E12.5) ICR mice (Harlan) were cultured on Nuclepore membranes (1 μ m pore size) in serum-free DMEM/F12 medium as described previously¹⁰.

Laser microdissection

SMGs were photographed 10 h after epithelial clefts were visible, washed with PBS, embedded in Tissue-Tek OCT compound (Sakura) in cryomoulds (Lab-Tek), and frozen in solid CO₂–methanol. Under RNase-free conditions, 12- μ m serial cryostat sections of SMGs were mounted on LMD foil slides (Leica). The unfixed sections were immediately stored at –80 °C. Toluidine blue staining was used to identify the SMG epithelium (www.arctur.com/resources/protocol_toluidine.htm). RNaseOUT ribonuclease inhibitor (400 U ml^{–1}; Invitrogen) was added to the toluidine blue solution. Toluidine blue-stained tissue sections were used for laser microdissection by an LMD system (Leica) with a pulsed ultraviolet laser on an upright automated microscope. About 500 cells from each type of region depicted in Fig. 1 were excised by the laser and transferred to caps of 0.5-ml PCR tubes (Eppendorf).

Quantitative real-time RT–PCR

Total RNA extraction, treatment with DNase I, and reverse transcription to synthesize complementary DNA were performed as described²⁸. Quantification of PCR products was performed by measuring fluorescence from the progressive binding of SYBR green I dye to double-stranded DNA using the ABI Prism 7700 sequence detection system (Applied Biosystems). The relative quantification value of PCR transcripts was calculated by using the manufacturer's protocol with normalization to glyceraldehyde-3-phosphate dehydrogenase (G3PDH). Fibronectin expression was calculated relative to the average of values from the mesenchymal transcripts. Data were collected from three cultures

containing six explants for each of the experimental groups and are expressed as means $\pm$ s.e.m. The sets of primers that we used were as follows: fibronectin (5'-TACCAAGGTCAATCCACACCCC-3' and 5'-CAGATGGCAAGAAAGCAGAGG-3'), ribosomal protein L19 (5'-AAATCGCCAATGCCAATCC-3' and 5'-TCCTCATCCTTCTCATCCAGGTC-3'), G3PDH (5'-CCATCACCATTCTCCAGGAG-3' and 5'-GCATGGACTGTGGTCATGAG-3') and E-cadherin (5'-ACGTATCAGGGTCAAGT GCC-3' and 5'-CCTGACCCACACCAAGTCT-3').

Whole-mount immunofluorescence microscopy

E12.5 SMGs cultured for 10 h were fixed in 4% paraformaldehyde and permeabilized with 0.1% Triton X-100 and then washed in PBS. The samples were incubated with M.O.M. blocking reagent (Vector) and then overnight with primary antibodies at 4 °C. After being rinsed with PBS containing 0.5% Tween 20 (PBST), the tissues were incubated with Cy3-labelled or Cy5-labelled donkey anti-mouse or donkey anti-rabbit IgG overnight at 4 °C in the presence of SYBR green I (dilution 1:500,000; Molecular Probes). The tissues were then washed in PBST and mounted on slides with Secure-Seal imaging spacers (Grace Bio-Labs) and examined by confocal microscopy (LSM 510; Zeiss).

Whole-mount fluorescence *in situ* hybridization

Digoxigenin-labelled antisense and sense RNA riboprobes contained a coding region segment of mouse fibronectin (GenBank accession number BC025521; residues 2466–2831). SMGs were fixed with 4% paraformaldehyde, treated with 10 μ g ml^{–1} proteinase K and incubated with RNA probe (500 ng ml^{–1}) at 55 °C for 24 h. For signal amplification, a rabbit anti-digoxigenin antibody conjugated with horseradish peroxidase was used to catalyse the deposition of biotin-tyramide (GenPoint kit; DAKO). Further amplification was achieved by adding rabbit anti-biotin conjugated with horseradish peroxidase (DAKO) and biotin-tyramide. The signal was detected with Cy5-labelled streptavidin (Jackson ImmunoResearch). SMGs were simultaneously immunostained with anti-perlecan antibody (Chemicon International) and donkey anti-mouse IgG tagged with Cy3 and were examined by confocal microscopy.

Inhibition by antibodies

SMGs were treated with a polyclonal rabbit antiserum against fibronectin (R745)²⁹ with a rabbit serum control or with rat monoclonal antibody 333 anti-mouse fibronectin with a rat IgG control (data not shown). All integrin antibodies were from BD Biosciences. SMGs were cultured for up to 48 h, and branching morphogenesis was monitored daily and photographed. Five SMGs were analysed for each concentration in dose–response experiments. All experiments were repeated at least three times; data are shown from a representative experiment.

siRNA and exogenous fibronectin

RNA interference was performed with siRNA as described²⁰. The fibronectin sequence for siRNA targeting, 5'-AACAAATCTCTGCTGGGAC-3', was 126–146 bases downstream of the first nucleotide of the start codon of mouse fibronectin cDNA. The order of fibronectin nucleotides was scrambled to generate the control sequence 5'-AACTCGACATAGCGACTCTG-3'. The siRNA oligonucleotides were synthesized, purified and duplexed by Dharmacon Research, which also provided the Scramble I Duplex negative control. SMGs were transfected with 500 nM siRNA duplex in regular serum-free culture medium by using Oligofectamine reagent (Invitrogen). Human plasma fibronectin was isolated as described³⁰ and dialysed extensively against culture medium before addition to SMG cultures.

HSG cell culture and cellular fibronectin treatment

Passage 42–46 HSG cells were maintained in DMEM and F12 medium (1:1) with 5% fetal bovine serum and antibiotics. Cellular fibronectin isolated as described¹⁷ was neutralized with 100 mM sodium phosphate pH 7.2 and left to form aggregates for 24 h at 4 °C. The cellular fibronectin was added at 25 μ g ml^{–1} to HSG cultures and incubated for 24 h. Cultures were fixed and immunostained as described²⁹, and examined by confocal microscopy. Similar results were obtained with different fixation and microscopy methods (see Supplementary Information 6).

Received 3 March; accepted 21 April 2003; doi:10.1038/nature01712.

- Gilbert, S. F. *Developmental Biology* 683–687 (Sinauer Associates, Sunderland, Massachusetts, 1997).
- Davies, J. A. Do different branching epithelia use a conserved developmental mechanism? *BioEssays* **24**, 937–948 (2002).
- Metzger, R. J. & Krasnow, M. A. Genetic control of branching morphogenesis. *Science* **284**, 1635–1639 (1999).
- Mosher, D. F. *Fibronectin* (Academic, San Diego, California, 1989).
- Hynes, R. O. *Fibronectins* (Springer, New York, 1990).
- Hogan, B. L. & Yingling, J. M. Epithelial/mesenchymal interactions and branching morphogenesis of the lung. *Curr. Opin. Genet. Dev.* **8**, 481–486 (1998).
- Grobstein, C. Morphogenetic interaction between embryonic mouse tissues separated by a membrane filter. *Nature* **172**, 869–871 (1953).
- Spooner, B. S., Thompson-Pletscher, H. A., Stokes, B. & Bassett, K. E. in *Developmental Biology* (ed. Steinberg, M. S.) vol. 3, 225–260 (Plenum, New York, 1985).
- Spooner, B. S. & Wessells, N. K. An analysis of salivary gland morphogenesis: Role of cytoplasmic microfilaments and microtubules. *Dev. Biol.* **27**, 38–54 (1972).
- Hoffman, M. P. *et al.* Gene expression profiles of mouse submandibular gland development: FGFR1 regulates branching morphogenesis *in vitro* through BMP- and FGF-dependant mechanisms. *Development* **129**, 5767–5778 (2002).
- Nogawa, H. & Takahashi, Y. Substitution for mesenchyme by basement-membrane-like substratum and epidermal growth factor in inducing branching morphogenesis of mouse salivary epithelium. *Development* **112**, 855–861 (1991).

12. Bernfield, M. R., Banerjee, S. D. & Cohn, R. H. Dependence of salivary epithelial morphology and branching morphogenesis upon acid mucopolysaccharide-protein (proteoglycan) at the epithelial surface. *J. Cell Biol.* **52**, 674–689 (1972).
13. Kadoya, Y. *et al.* Antibodies against domain E3 of laminin-1 and integrin α_6 subunit perturb branching epithelial morphogenesis of submandibular gland, but by different modes. *J. Cell Biol.* **129**, 521–534 (1995).
14. Hay, E. D. *Cell Biology of Extracellular Matrix* (Plenum, New York, 1991).
15. Nakanishi, Y., Nogawa, H., Hashimoto, Y., Kishi, J. & Hayakawa, T. Accumulation of collagen III at the cleft points of developing mouse submandibular epithelium. *Development* **104**, 51–59 (1988).
16. Velling, T., Risteli, J., Wennerberg, K., Mosher, D. F. & Johansson, S. Polymerization of type I and III collagens is dependent on fibronectin and enhanced by integrins $\alpha_1\beta_1$ and $\alpha_3\beta_1$. *J. Biol. Chem.* **277**, 37377–37381 (2002).
17. Yamada, K. M. & Olden, K. Fibronectins—adhesive glycoproteins of cell surface and blood. *Nature* **275**, 179–184 (1978).
18. Hardman, P. & Spooner, B. S. Localization of extracellular matrix components in developing mouse salivary glands by confocal microscopy. *Anat. Rec.* **234**, 452–459 (1992).
19. Menko, A. S., Zhang, L., Schiano, F., Kreidberg, J. A. & Kukuruzinska, M. A. Regulation of cadherin junctions during mouse submandibular gland development. *Dev. Dyn.* **224**, 321–333 (2002).
20. Elbashir, S. M., Harborth, J., Weber, K. & Tuschl, T. Analysis of gene function in somatic mammalian cells using small interfering RNAs. *Methods* **26**, 199–213 (2002).
21. Nakanishi, Y., Sugiura, F., Kishi, J. & Hayakawa, T. Collagenase inhibitor stimulates cleft formation during early morphogenesis of mouse salivary gland. *Dev. Biol.* **113**, 201–206 (1986).
22. Kashimata, M. *et al.* The ERK-1/2 signaling pathway is involved in the stimulation of branching morphogenesis of fetal mouse submandibular glands by EGF. *Dev. Biol.* **220**, 183–196 (2000).
23. Roman, J. Fibronectin and fibronectin receptors in lung development. *Exp. Lung Res.* **23**, 147–159 (1997).
24. Jiang, S. T., Chuang, W. J. & Tang, M. J. Role of fibronectin deposition in branching morphogenesis of Madin–Darby canine kidney cells. *Kidney Int.* **57**, 1860–1867 (2000).
25. Shirasuna, K., Sato, M. & Miyazaki, T. A neoplastic epithelial duct cell line established from an irradiated human salivary gland. *Cancer* **48**, 745–752 (1981).
26. Levenberg, S., Katz, B. Z., Yamada, K. M. & Geiger, B. Long-range and selective autoregulation of cell–cell or cell–matrix adhesions by cadherin or integrin ligands. *J. Cell Sci.* **111**, 347–357 (1998).
27. Grobstein, C. & Cohen, J. Collagenase: Effect on the morphogenesis of embryonic salivary epithelium *in vitro*. *Science* **150**, 626–628 (1965).
28. Sakai, T., Larsen, M. & Yamada, K. M. *Current Protocols in Cell Biology* (eds Bonifacino, J. S. *et al.*) 19.3.1–19.3.30 (Wiley, New York, 2002).
29. Cukierman, E., Pankov, R., Stevens, D. R. & Yamada, K. M. Taking cell–matrix adhesions to the third dimension. *Science* **294**, 1708–1712 (2001).
30. Miekka, S. L., Ingham, K. C. & Menache, D. Rapid methods for isolation of human plasma fibronectin. *Thromb. Res.* **27**, 1–14 (1982).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements T.S., M.L. and K.M.Y. conceived and performed the experiments; T.S. and M.L. designed the data analysis and T.S. performed it; T.S. and K.M.Y. co-wrote the paper. We thank M. P. Hoffman and R. Pankov for advice. T.S. was supported by a fellowship from the Japan Society for the Promotion of Science, and M.L. was supported by a fellowship from the NIH.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.M.Y. (kenneth.yamada@nih.gov).

Regulation of flowering time by light quality

Pablo D. Cerdán & Joanne Chory

Howard Hughes Medical Institute and Plant Biology Laboratory, The Salk Institute for Biological Studies, La Jolla, California 92037, USA

The transition to flowering in plants is regulated by environmental factors such as temperature and light¹. Plants grown under dense canopies or at high density perceive a decrease in the ratio of red to far-red incoming light. This change in light quality serves as a warning of competition, triggering a series of responses known collectively as the ‘shade-avoidance syndrome’. During shade avoidance, stems elongate at the expense of leaf expansion, and flowering is accelerated^{2,3}. Of the five phytochromes—a family of red/far-red light photoreceptors—in *Arabidopsis*, phytochrome B (phyB) has the most significant role in shade-avoidance responses^{4,5}, but the mechanisms by which phyB

regulates flowering in response to altered ratios of red to far-red light are largely unknown. Here we identify PFT1 (PHYTOCHROME AND FLOWERING TIME 1), a nuclear protein that acts in a phyB pathway and induces flowering in response to suboptimal light conditions. PFT1 functions downstream of phyB to regulate the expression of *FLOWERING LOCUS T* (*FT*), providing evidence for the existence of a light-quality pathway that regulates flowering time in plants.

The phytochromes and the blue/ultraviolet-A photoreceptors called cryptochromes (cry1 and cry2 in *Arabidopsis*) are critical photoreceptors that regulate seedling emergence and floral induction^{6,7}. Several components involved in phytochrome signalling in seedlings have been isolated and characterized in recent years⁷. Seedlings defective in phyA signalling are tall under far-red light (FR), whereas those defective in phyB signalling are tall under red light (R). Later in development, light perceived by phyA and cry2 generates a signal that is required by CO (CONSTANS) to activate *FT*, a mechanism essential for day-length discrimination⁸. Despite the large number of phytochrome components identified, the mechanisms by which phytochromes regulate seedling development and flowering time are still largely unknown^{7,9}. To identify components involved in phytochrome regulation of flowering time in *Arabidopsis*, we devised a new genetic screen (see Methods). Seedlings showing an enhanced response to pulses of red light were selected and transplanted to soil and then scored for defects in flowering time. One of the recessive mutants, *pft1*, is characterized in detail here.

Consistent with a role in phytochrome signalling, *pft1* mutants displayed small but significant effects on hypocotyl-length inhibition under both red and far-red light (Fig. 1a). The mutants were hypo-responsive to far-red and hyper-responsive to red. These altered responses to light required functional phyA and phyB (Fig. 1a). The levels of phyA and phyB protein were unaltered in *pft1* compared with wild type, suggesting that PFT1 affects processes downstream of phyA and phyB (data not shown). Negative effects of phyA signalling on phyB signalling, as well as negative effects of phyB on phyA function, have been reported previously^{10,11}. The *pft1* mutation has opposite effects on phyA- and phyB-mediated responses, indicating that PFT1 may function in phytochrome signalling at a node where antagonistic interactions occur between phyA and phyB. The *pft1* mutants had normal responses to blue light (data not shown), suggesting that PFT1 is specific for phytochrome action and is not a general regulator of light responsiveness.

pft1 mutants displayed a late-flowering phenotype when grown under long-day conditions (LD, 16 h light/8 h dark; Fig. 1b, c). phyA and phyB regulate flowering time in opposing ways: phyB acts to delay flowering under both long- and short-day (SD, 9 h light/15 h dark) conditions, whereas phyA weakly promotes flowering in long days⁶. We measured flowering time under both LD and SD conditions in *pft1*, *pft1 phyB*, *pft1 phyA* and *pft1 phyA phyB* mutants, and compared the results with *phyB*, *phyA* and *phyA phyB* mutants (Fig. 1c, d). *pft1* was late flowering and completely suppressed the early-flowering phenotype of *phyB* in both LD and SD conditions, strongly suggesting that PFT1 is essential for phyB regulation of flowering time. The effects of *pft1* on the photoperiodic induction of flowering were relatively minor in both wild-type and in phytochrome mutant backgrounds (Fig. 1c, d), indicating that PFT1 is unlikely to act in the photoperiod pathway that induces flowering.

PFT1 may also act downstream of phyA, as the effects of *phyA* and *pft1* mutations in delaying flowering under LD were not additive. However, a phyA-dependent photoperiod response is still present in the absence of PFT1 or phyB, because *pft1 phyA phyB* mutants flowered significantly later than *pft1 phyB* mutants under LD but not SD. It has been proposed¹² that phyA can promote flowering by two independent mechanisms: by suppressing phyB function, and independently of phyB. Our data are consistent with this report, and they suggest that the effects of *pft1* on phyA regulation of flowering

time may be an indirect effect of changes in the *phyB* pathway.

The increased petiole length of *phyB* mutants was largely unaffected in the *pft1 phyB* double mutants (Fig. 1e). These results, together with the mild effects of *pft1* on hypocotyl length, the strong response to photoperiod in *pft1* mutants in phytochrome-deficient backgrounds and the complete suppression of *phyB*'s early-flowering phenotype suggest that the main role of PFT1 in phytochrome signalling is to regulate flowering time downstream of *phyB* in a photoperiod-independent pathway.

Cloning of *PFT1* revealed that it encodes a putative protein of 836 amino acids, with a predicted vWF-A (von Willebrand factor type A) domain in the amino terminus^{13,14} and a glutamine-rich region in the carboxy terminus, reminiscent of some transcriptional activators¹⁵ (Fig. 2a). vWF-A domains are widely distributed among all phyla¹⁴. They are involved in various cellular processes, and a high proportion have a divalent cation-binding site that in some cases has been shown to mediate protein–protein interactions¹⁶. The DxSxS motif involved in coordination with a divalent cation is converted to ExTxA in PFT1. It is unclear whether this partially conserved motif in PFT1 can still bind a metal.

The *pft1* mutant produced a truncated messenger RNA (Fig. 2b), and transformation with a genomic copy of *PFT1* rescued both the seedling and flowering time defects (Fig. 2d–f). We found that *PFT1* is limiting for flowering under SD, as overexpression of *PFT1* caused an early-flowering phenotype under these conditions (Fig. 2f, g).

This result is also consistent with the strongest effects of the *phyB* mutation observed under SD (Fig. 1c, d). Transformation of *pft1* plants with *PFT1* complementary DNA under the strong 35S promoter also produced a high proportion of co-suppressed plants whose phenotypes were not more severe than the *pft1* mutant. These data support the interpretation that the *pft1* allele is a strong or null allele (Fig. 2f, g and data not shown).

To study the subcellular localization of PFT1, we fused the gene encoding green fluorescent protein (GFP) to either the N or the C terminus of *PFT1* in the context of the genomic clone. These constructs encoded a functional protein, as they complemented the *pft1* phenotype (data not shown). Both N- and C-terminal GFP fusions localized to the nucleus (Fig. 2c and data not shown). The subcellular location of PFT1–GFP did not change in response to red, far-red or white light (data not shown). These observations, together with the mild effect of the *phyB* mutation on *PFT1* mRNA levels (see Supplementary Information), indicate that *phyB* regulation of *PFT1* activity occurs post-transcriptionally or through association with other proteins.

The glutamine-rich region of PFT1 (Fig. 2a), its nuclear localization (Fig. 2c) and its ability to activate transcription in yeast when fused to the LexA DNA-binding domain (data not shown) suggested that PFT1 might function as a transcriptional co-activator. To test the nature of *phyB* suppression by *pft1*, we investigated its role in the regulation of key genes that affect flowering time. *FT* is an

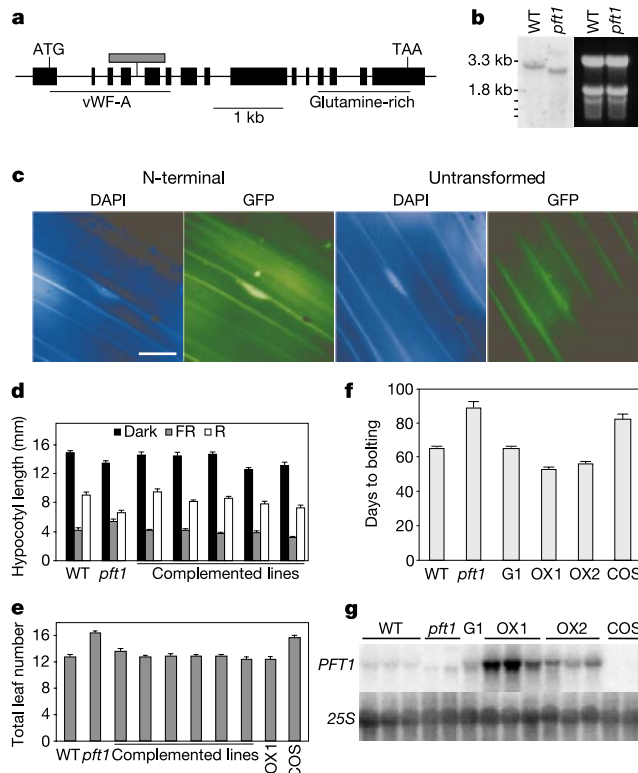
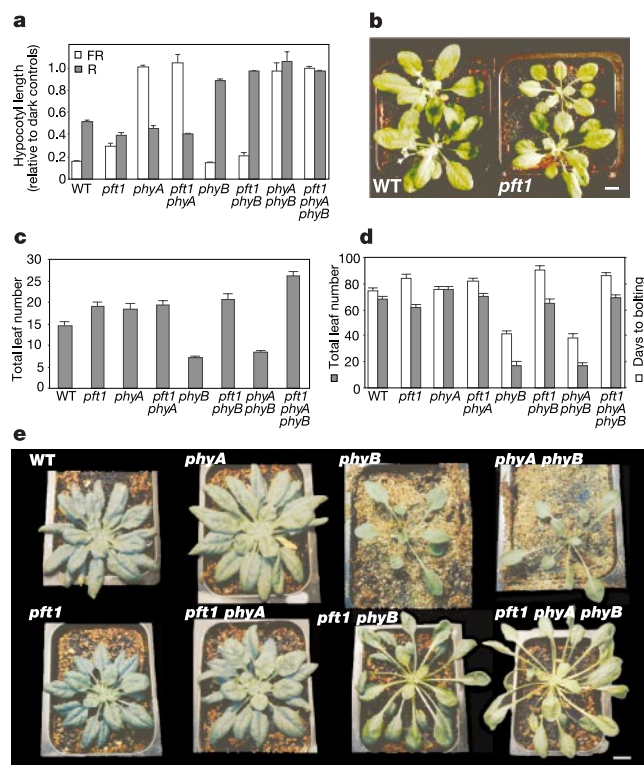


Figure 1 Phenotypes of *pft1*, *phyA* and *phyB* single, double and triple mutants.

a, Hypocotyl length of 5-day-old seedlings of the indicated genotypes grown for 4 days under 3 $\mu\text{mol m}^{-2} \text{s}^{-1}$ of far-red light (FR) or 10 $\mu\text{mol m}^{-2} \text{s}^{-1}$ of red light (R). **b**, Wild-type (WT) and *pft1* plants grown as described in **c**. **c**, **d** Flowering time of the indicated genotypes grown under long-day (16 h white light; 8 h dark; **c**) and short-day (9 h white light; 15 h dark; **d**) conditions. **e**, The genotypes used in **a**, **c** and **d**, cultivated in short days for 5 (*phyB* and *phyA phyB*) or 9 (WT, *pft1*, *phyA*, *pft1 phyA*, *pft1 phyB* and *pft1 phyA phyB*) weeks. Note that in *phyB* and *phyA phyB*, the apical meristem is already converted to an inflorescence meristem, whereas no sign of change is observed in the *pft1* background. Data represent averages of three independent experiments (**a**) and eight to ten plants (**c**, **d**) $\pm$ s.e. Scale bar, 1 cm.

Figure 2 Molecular characterization of *PFT1*. **a**, Structure of the *PFT1* gene. Black boxes represent exons; lines represent introns. Predicted domains and the T-DNA insertion (out of scale) are indicated. Scale bar, 1 kb. **b**, Northern blot of WT and *pft1* (left) and the corresponding gel (right). **c**, Localization of GFP–PFT1 fusions and the corresponding controls, 4,6-diamidino-2-phenylindole (DAPI) staining and non-transgenic lines, as indicated. Scale bar, 10 μm . **d**, **e**, **f** Molecular complementation of the hypocotyl-length phenotype (**d**), late-flowering phenotype in long days (**e**) and delayed bolting in short days (**f**). Different transgenic lines for the *PFT1* gene were used, as indicated. OX1 and OX2 are 35S–PFT1-overexpressing lines in *pft1* and WT backgrounds, respectively. COS is a co-suppressed line in *pft1* background, and G1 is a complemented line. **g**, *PFT1* mRNA levels and the corresponding ribosomal RNA control for the lines shown in **f**.

integrator of several flowering time pathways^{8,17–19}. *FT* expression is lower in *phyA* and *cry2* mutants in conditions where these mutants are late flowering⁸. We analysed the expression patterns of *FT* mRNA in wild-type, *phyB*, *pft1* and *phyB pft1* double mutants. We first tested expression levels in 8-day-old seedlings, as previously described^{8,20,21}. In LD-grown seedlings, *FT* expression was substantially higher in *phyB* mutants than in wild type (Fig. 3a, b), indicating that *phyA* and *phyB* have opposing roles in flowering by differentially regulating *FT* expression. By contrast, these differences were not found in 8-day-old seedlings grown in SD, consistent with previous reports²¹ (data not shown). This could be due to the lack of commitment to flowering of SD-grown seedlings at this early stage of development (wild-type plants produce about 70 leaves in SD conditions, whereas primary leaves are just emerging in 8-day-old seedlings). Thus, we tested *FT* mRNA levels in 26-day-old plants grown in SD, a time at which *phyB* mutants are roughly two weeks from showing visible floral buds. Under these conditions, we observed a large increase in *FT* mRNA levels in *phyB* mutants compared with wild-type plants (Fig. 3c, d). These data thus

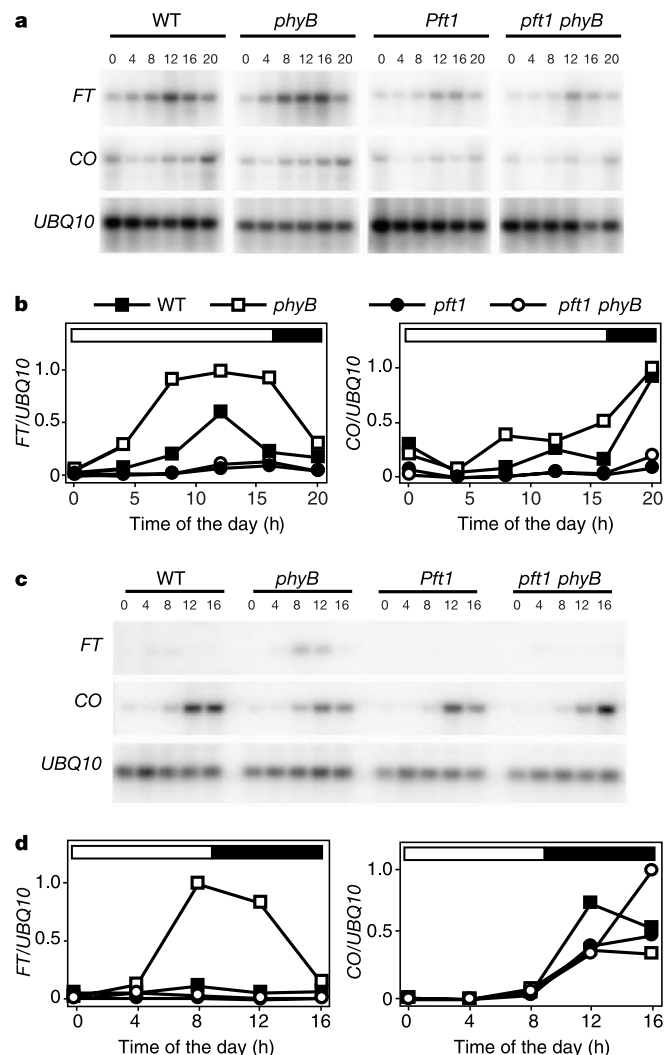


Figure 3 *FT* and *CO* mRNA levels in WT, *phyB*, *pft1* and *pft1 phyB* mutants. **a**, *FT* and *CO* expression in 8-day-old seedlings grown under long-day conditions and harvested at the indicated times. 0, lights on. **b**, Quantification of data shown in **a** relative to the *UBQ10* control. **c**, *FT* and *CO* expression in 26-day-old plants grown under short-day conditions and harvested at the indicated times. 0, lights on. **d**, Quantification of data shown in **c** relative to the *UBQ10* control. Data in **a** and **c** represent two independent experiments.

provide a molecular mechanism for the early flowering time of *phyB* mutants in both LD and SD. Moreover, *FT* mRNA levels were low in *pft1* and *pft1 phyB* double mutants in all the conditions tested (Fig. 3a–d), pointing to the molecular mechanism for the suppression of the early-flowering-time phenotype of *phyB* mutants by *pft1*.

CO is a key component²² of the photoperiod pathway, which integrates information from the circadian clock²⁰ as well as *phyA* and *cry2* signalling⁸. CO also directly activates *FT* expression²³. We found higher levels of *CO* mRNA in LD-grown *phyB* seedlings than in wild type, and lower levels in the *pft1* background, suggesting that PFT1 might function upstream of CO in the regulation of *FT* (Fig. 3a, b). However, in SD-grown plants, we were unable to detect increased levels of *CO* in *phyB* mutants, nor were the levels decreased in the *pft1* background (Fig. 3c, d). We extended this analysis to *SOC1*, another floral integrator^{24,25}, and did not observe a correlation of *SOC1* mRNA levels with flowering time, in either *phyB* or *pft1* mutant backgrounds (see supplement to Fig. 3 in Supplementary Information). The lack of a significant correlation between *CO* and *SOC1* mRNA levels with *phyB* and *pft1* flowering times suggests that *phyB* regulates *FT* mRNA levels by a PFT1-dependent mechanism that does not involve changes in *CO* or *SOC1* mRNA levels. It has been suggested that light post-translationally regulates CO activity²⁰. However, *pft1* does not have a significant effect on flowering time or on *FT* mRNA levels in a CO-overexpressing line, arguing that PFT1 does not have a significant role in the post-translational regulation of CO (Fig. 4a and data not shown).

If PFT1 has a specific role in a *phyB* pathway that regulates flowering time in response to changes in light quality, then *pft1* mutants should be unable to accelerate flowering in response to shade conditions. We grew wild-type plants and *pft1* mutants in SD

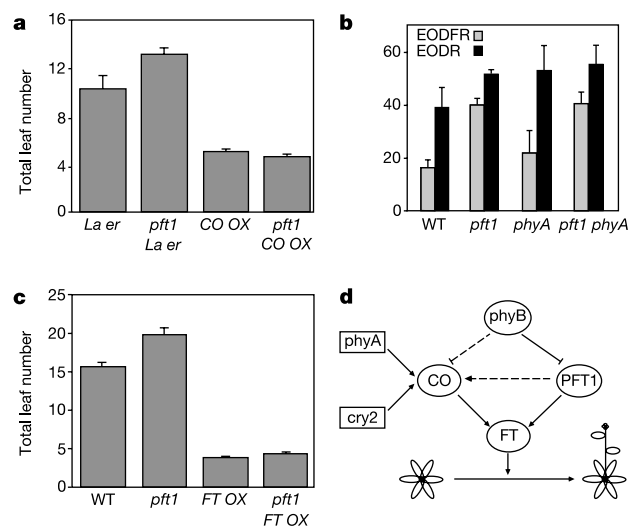


Figure 4 PFT1 acts in a light-quality pathway upstream of *FT*. **a**, Flowering time of WT (cv. *Landsberg erecta*), *pft1*, *CO* overexpressor in WT (*CO OX*) and *pft1* (*pft1 CO OX*) backgrounds. For this particular experiment the *pft1* mutation was introgressed four times into the *L. erecta* ecotype before crossing to *CO OX* lines. **b**, Flowering time of the indicated genotypes grown under short-day conditions in factorial combination with 15 min of red or far-red light at the end of each photoperiod (EODR and EODFR). Data are mean $\pm$ s.e. of seven to nine plants. **c**, Flowering time of WT, *pft1*, *FT*-overexpressor in WT (*FT OX*) and *pft1* (*pft1 FT OX*) backgrounds. Data from **a** and **c** are mean $\pm$ s.e. of at least ten plants grown under long-day conditions. **d**, A proposed model to explain PFT1 involvement in a light-quality pathway that regulates flowering time. With the exception of the direct regulation of *FT* by CO (ref. 23), other arrows do not represent direct regulation or interaction. *phyA* and *cry2* seem to mediate a direct effect of light on CO activity⁸. The role of *FT* in the proposed pathway is based on previous studies^{18,19}. The placement of PFT1 is deduced from the work presented here.

followed by an end-of-day pulse of far-red (EODFR) or red (EODR) light. The EODFR pulse mimics the effects of shade, including accelerated flowering. The promotion of flowering by the EODFR treatment (compared with EODR controls) was severely impaired in the *pft1* and *pft1 phyA* mutants compared with wild type and *phyA*, respectively (Fig. 4b), consistent with a role in signalling after perception of low red/far-red ratios. It is possible that PFT1 acts in an autonomous pathway, independently of changes in light quality, but we feel that this is unlikely, as *pft1* effects on flowering are mild, except in a *phyB* background, and the *pft1* mutation specifically impairs the response to low red/far-red ratios. As mentioned, it is also unlikely that PFT1 has a role in a photoperiod pathway, as *pft1* and *pft1 phyB* double mutants show a strong response to photoperiod (Fig. 1c, d), in contrast to *co* and *co phyB* double mutants^{3,22,26}. This notion is reinforced by our observation that PFT1 can regulate *FT* mRNA levels by mechanisms that do not involve CO. Finally, the pattern of expression of *FT* in *phyB* and *pft1* single and double mutants indicates that the activation of *FT* occurs downstream of PFT1 in a *phyB* pathway, consistent with the lack of effect of the *pft1* mutation in a *FT*-overexpressing background (Fig. 4c).

A 'light-quality' pathway has been proposed⁹ to explain the role of *phyB* and other stable phytochromes (*phyD* and *phyE* in *Arabidopsis*) in flowering. This light-quality perception pathway may be involved in the shade-avoidance syndrome triggered in response to low red/far-red ratios^{2,3}. Here we present genetic and molecular evidence that this light-quality pathway exists, and we incorporate a new component, PFT1, into this signalling pathway. Our results are consistent with a model (Fig. 4d) in which PFT1 acts in a light-quality pathway downstream of *phyB*. The resulting signal is then integrated with the photoperiod pathway through the modulation of *FT* transcription. □

Methods

Plant material and growth conditions

Except when stated otherwise, all experiments were done in the Columbia accession. The alleles for the different mutants used were *phyA211* and *phyB9*. *FT*- and *CO*-overexpressing lines were provided by D. Weigel. Seeds were sterilized using chlorine in a vapour phase. For RNA extraction, we used total seedlings grown in MS media without sucrose, or shoots from mature plants grown on soil. For hypocotyl measurements, seedlings were grown in water–agar medium. Red and far-red light treatments were done in chambers of light-emitting diodes (Percival Scientific) at 23 °C. White light sources were always from fluorescent tubes, 40–70 $\mu\text{mol m}^{-2}\text{s}^{-1}$ for LD and SD, respectively, except for incandescent light extensions (LI) with 2 $\mu\text{mol m}^{-2}\text{s}^{-1}$ of photosynthetic active radiation (400–700 nm).

For the genetic screen, activation-tagged transfer DNA lines were obtained from the Arabidopsis Stock Center as pools representing around 100 lines each. These lines are at least in the T3 generation, so recessive mutations can also be found. Each pool was plated on water–agar medium, and stratified for 3 days. Germination was induced by 1 h of white light, and then seedlings received 5 min of red light every 24 h. On the fourth day, seedlings were scored for signs of de-etiolation, such as shorter hypocotyls or partially opened cotyledons, and transplanted to soil to score adult phenotypes.

Molecular characterization of PFT1

Genomic DNA was extracted from *pft1* mutants, cut with *HindIII* and *EcoRI* for right border rescue, and with *BamHI* and *SpeI* for left border rescue, and ligated and electroporated into SURE competent cells (Stratagene). Genomic DNA was successfully rescued with *HindIII*, *BamHI* and *SpeI* restriction enzymes. We designed specific oligonucleotides to detect by polymerase chain reaction (PCR) the T-DNA insertion in hemizygous and homozygous lines. The following three primers were used in PCR: 5'-CAGAGGAACCTGTTTCTACTGTGAGCT-3', 5'-CGTTACTTGTTGAGCTTGGCTGAAGGA-3' and 5'-TCCCGGACATGAAGCCATTATATGTA-3'. The expected PCR products were 563 base pairs (bp) for the wild type and 491 bp for the mutant, and both bands were detected in hemizygotes.

The T-DNA co-segregated with the *pft1* mutation in 140 chromosomes analysed. We also tested 36 F3 populations and noted that the Basta resistance (conferred by the T-DNA) also co-segregated with the *pft1* mutation. Southern blot data and the rescued genomic DNA sequence agreed with a single insertion on BAC F2J7 from chromosome I.

Intron–exon junctions were derived by comparison with EST clone APZL03h11R (GenBank accession number AV528220). To confirm the 5' end, RACE (rapid amplification of cDNA ends)–PCR was performed using the GeneRacer kit according to the manufacturer's instructions (Invitrogen). In *pft1* mutants, an mRNA species of lower molecular weight was observed (Fig. 3b). We used RACE–PCR to characterize this form

further, and found that it corresponds to a 2.5-kilobase (kb) truncated form of *PFT1* that initiates within the T-DNA.

The genomic *PFT1* clone was subcloned from BAC F2J7 in two consecutive steps into binary plasmid pPZP212, finally as a *PstI*–*SacI* 8.5-kb fragment and introduced into *pft1* mutants by transformation with *Agrobacterium tumefaciens*. Transgenic lines were selected on MS media supplemented with 50 $\mu\text{g ml}^{-1}$ of kanamycin, screened for single-locus insertions in the T2 generation, and the homozygous lines from the T3 generation used for physiological experiments.

To make the GFP fusion proteins, *EcoRI* sites were introduced in-frame right before the ATG start codon (for GFP–PFT1 fusions) and right before the TAA stop codon (for PFT1–GFP fusions). The GFP-coding region amplified by PCR was subcloned into these *EcoRI* sites, and final constructs were confirmed by sequencing before being used for plant transformation as described above.

To assess a phenotype of overexpression, the *PFT1* cDNA was subcloned in two steps as a *BamHI*–*PstI*–*PstI* fragment into CHF1 and CHF3 plasmids under the 35S constitutive promoter. After plant transformation, lines were selected on kanamycin (CHF3-derived plasmids) or gentamycin (70 $\mu\text{g ml}^{-1}$ CHF1-derived plasmids).

mRNA quantification

Specific mRNAs were extracted and quantified by RT–PCR (PCR with reverse transcription), essentially as described²¹. mRNA was reverse transcribed using the 1st cDNA synthesis kit according to the manufacturer's instructions (Invitrogen). Primers, annealing temperatures and the size of PCR products expected were as follows: *FT*, 5'-GCTACAACTGGAACAACCTTTGGCAAT-3', 5'-TATAGGCATCATCACCGTT CGTTACTC-3', 63 °C, 365 bp; *CO*, 5'-AAACTCTTTTCAGCTCCATGACCACTACT-3', 5'-CCATGGATGAAATGATGCGTTATGGTTA-3', 62 °C, 453 bp; *UBQ10*, 5'-GGTGT CAGAACTCTCCACCTCAAGAGTA-3', 5'-TCAATTCTCTACCGTGATCAAGAT GCA-3', 64 °C, 318 bp; *TSF*, 5'-GAAATTCTACACCTTGGTTATGGTGGA-3', 5'-CTTTACATCGATGTCAGCATATGCATCA-3', 63 °C, 484 bp. In all cases, at least one of the primers used for PCR spanned intron–exon junctions, and amplification of genomic DNA was undetectable in non-retrotranscribed controls. *SOCI/AGL20* was quantified by RT–PCR, as described²⁷. PCR products were detected by Southern blot using standard methodology, and quantified using a Phosphorimager (Molecular Dynamics) in the exponential range of amplification.

Received 31 October 2002; accepted 14 April 2003; doi:10.1038/nature01636.

- Levy, Y. Y. & Dean, C. Control of flowering time. *Curr. Opin. Plant Biol.* **1**, 49–54 (1998).
- Ballare, C. L. Keeping up with the neighbours: phytochrome sensing and other signalling mechanisms. *Trends Plant Sci.* **4**, 201 (1999).
- Halliday, K. J., Koornneef, M. & Whitelam, G. C. Phytochrome B and at least one other phytochrome mediate the accelerated flowering response of *Arabidopsis thaliana* L. to low red/far-red ratio. *Plant Physiol.* **104**, 1311–1315 (1994).
- Aukerman, M. J. *et al.* A deletion in the *PHYD* gene of the *Arabidopsis* Wassilewskija ecotype defines a role for phytochrome D in red/far-red light sensing. *Plant Cell* **9**, 1317–1326 (1997).
- Devlin, P. F., Patel, S. R. & Whitelam, G. C. Phytochrome E influences internode elongation and flowering time in *Arabidopsis*. *Plant Cell* **10**, 1479–1487 (1998).
- Lin, C. Photoreceptors and regulation of flowering time. *Plant Physiol.* **123**, 39–50 (2000).
- Quail, P. H. Phytochrome photosensory signalling networks. *Nature. Rev. Mol. Cell Biol.* **3**, 85–93 (2002).
- Yanovsky, M. J. & Kay, S. A. Molecular basis of seasonal time measurement in *Arabidopsis*. *Nature* **419**, 388–392 (2002).
- Simpson, G. G. & Dean, C. *Arabidopsis*, the Rosetta stone of flowering time? *Science* **296**, 285–289 (2002).
- Cerdan, P. D. *et al.* Regulation of phytochrome B signaling by phytochrome A and PHY1 in *Arabidopsis thaliana*. *Plant J.* **18**, 499–507 (1999).
- Hennig, L., Poppe, C., Sweere, U., Martin, A. & Schafer, E. Negative interference of endogenous phytochrome B with phytochrome A function in *Arabidopsis*. *Plant Physiol.* **125**, 1036–1044 (2001).
- Mockler, T. *et al.* Regulation of photoperiodic flowering by *Arabidopsis* photoreceptors. *Proc. Natl Acad. Sci. USA* **100**, 2140–2145 (2003).
- Ponting, C. P., Schultz, J., Milpetz, F. & Bork, P. SMART: identification and annotation of domains from signalling and extracellular protein sequences. *Nucleic Acids Res.* **27**, 229–232 (1999).
- Ponting, C. P., Aravind, L., Schultz, J., Bork, P. & Koonin, E. V. Eukaryotic signalling domain homologues in archaea and bacteria. Ancient ancestry and horizontal gene transfer. *J. Mol. Biol.* **289**, 729–745 (1999).
- Escher, D., Bodmer-Glavas, M., Barberis, A. & Schaffner, W. Conservation of glutamine-rich transactivation function between yeast and humans. *Mol. Cell Biol.* **20**, 2774–2782 (2000).
- Hinshelwood, J. & Perkins, S. J. Metal-dependent conformational changes in a recombinant vWF-A domain from human factor B: a solution study by circular dichroism, fourier transform infrared and ¹H NMR spectroscopy. *J. Mol. Biol.* **298**, 135–147 (2000).
- Levy, Y. Y., Mesnage, S., Mylne, J. S., Gendall, A. R. & Dean, C. Multiple roles of *Arabidopsis* *VRN1* in vernalization and flowering time control. *Science* **297**, 243–246 (2002).
- Kardailsky, I. *et al.* Activation tagging of the floral inducer *FT*. *Science* **286**, 1962–1965 (1999).
- Kobayashi, Y., Kaya, H., Goto, K., Iwabuchi, M. & Araki, T. A pair of related genes with antagonistic roles in mediating flowering signals. *Science* **286**, 1960–1962 (1999).
- Suarez-Lopez, P. *et al.* *CONSTANS* mediates between the circadian clock and the control of flowering in *Arabidopsis*. *Nature* **410**, 1116–1120 (2001).
- Blazquez, M. A. & Weigel, D. Independent regulation of flowering by phytochrome B and gibberellins in *Arabidopsis*. *Plant Physiol.* **120**, 1025–1032 (1999).
- Putterill, J., Robson, F., Lee, K., Simon, R. & Coupland, G. The *CONSTANS* gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. *Cell* **80**, 847–857 (1995).
- Samach, A. *et al.* Distinct roles of *CONSTANS* target genes in reproductive development of *Arabidopsis*. *Science* **288**, 1613–1616 (2000).
- Hepworth, S. R., Valverde, F., Ravenscroft, D., Mouradov, A. & Coupland, G. Antagonistic regulation

of flowering-time gene *SOC1* by *CONSTANS* and *FLC* via separate promoter motifs. *EMBO J.* **21**, 4327–4337 (2002).

25. Lee, H. *et al.* The AGAMOUS-LIKE 20 MADS domain protein integrates floral inductive pathways in *Arabidopsis*. *Genes Dev.* **14**, 2366–2376 (2000).
26. Koornneef, M., Hanhart, C., Van Loenen-Martinet, P. & Blankestijn-de Vries, H. The effect of daylength on the transition to flowering in phytochrome-deficient, late-flowering and double mutants of *Arabidopsis thaliana*. *Physiol. Plant* **95**, 260–266 (1995).
27. Blazquez, M. A., Trenor, M. & Weigel, D. Independent control of gibberellin biosynthesis and flowering time by the circadian clock in *Arabidopsis*. *Plant Physiol.* **130**, 1770–1775 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank the Kasuza Research Institute and the *Arabidopsis* Stock Center for providing EST clone APZL03h11R and BAC F2J7, respectively; D. Weigel for providing *FT*- and *CO*-overexpressing lines, and for useful advice; P. Wigge, S. Mora-Garcia, D. Weigel, J. Nemhauser, M. Yanovsky and J. Casal for comments on the manuscript; L. Barden for figure preparation; and B. Smoot, N. Ga and M. Ledgerwood for technical assistance. This work was supported by a grant from the National Institutes of Health to J.C. (GM52413) and by the Howard Hughes Medical Institute. P.C. was a fellow of the PEW Latin American Fellows Program and Fundación Antorchas.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.C. (chory@salk.edu). The GenBank accession code for *PFT1* is AY170377.

CUL-4 ubiquitin ligase maintains genome stability by restraining DNA-replication licensing

Weiwei Zhong, Hui Feng, Fernando E. Santiago & Edward T. Kipreos

Department of Cellular Biology, University of Georgia, Athens, Georgia 30602, USA

To maintain genome stability, DNA replication is strictly regulated to occur only once per cell cycle. In eukaryotes, the presence of ‘licensing proteins’ at replication origins during the G1 cell-cycle phase allows the formation of the pre-replicative complex¹. The removal of licensing proteins from chromatin during the S phase ensures that origins fire only once per cell cycle¹. Here we show that the CUL-4 ubiquitin ligase temporally restricts DNA-replication licensing in *Caenorhabditis elegans*. Inactivation of CUL-4 causes massive DNA re-replication, producing cells with up to 100C DNA content. The *C. elegans* orthologue of the replication-licensing factor Cdt1 (refs 2, 3) is required for DNA replication. *C. elegans* CDT-1 is present in G1-phase nuclei but disappears as cells enter S phase. In cells lacking CUL-4, CDT-1 levels fail to decrease during S phase and instead remain constant in the re-replicating cells. Removal of one genomic copy of *cdt-1* suppresses the *cul-4* re-replication phenotype. We propose that CUL-4 prevents aberrant re-initiation of DNA replication, at least in part, by facilitating the degradation of CDT-1.

CUL-4 is a member of the cullin ubiquitin-ligase family⁴. Cullins function in cullin/RING finger complexes to catalyse the covalent attachment of ubiquitin to protein substrates to mark them for proteolysis⁵. In *C. elegans*, the *cul-4* gene is expressed throughout development. In adult hermaphrodites, *cul-4* messenger RNA is observed primarily in the germ line, with transient expression in the intestine of young adults (Fig. 1a, b; data not shown). *cul-4* maternal mRNA is present in early embryos and decreases during embryogenesis (Fig. 1c, d). In larvae, *cul-4* is broadly expressed, with high levels in proliferating tissues, notably in the intestine and germ line (Fig. 1e).

To probe *cul-4* function, we used RNA-mediated interference

(RNAi)⁶ to inactivate the *cul-4* gene. *cul-4* RNAi reduced *cul-4* mRNA to levels not significantly higher than background (Fig. 1f–h). The predominant *cul-4* RNAi phenotype was a developmental arrest at the L2 larval stage. We observed a pronounced increase in the size of blast cell nuclei in *cul-4* RNAi L2 larvae. In a *clr-1(e1745)* genetic background that allows the visualization of cell boundaries⁷, *cul-4* RNAi produced a marked increase in the size of seam cells (Fig. 2a, b). However, the size of nuclei in the non-proliferative hyp7 syncytial cell⁸ was unaffected by inactivation of *cul-4* (Fig. 2a, b). Larger cells were also observed among other *cul-4* RNAi blast cell lineages, including the M, Q, P and somatic gonad cell lineages (Fig. 2c–f; data not shown). Germ cells did not become enlarged. However, germ cells often entered meiosis prematurely to become sperm in the L2 larval stage; this is in contrast to the wild type, in which spermatogenesis occurs in the L4 larval stage⁹ (Fig. 2e, f). We assume that this premature spermatogenesis results

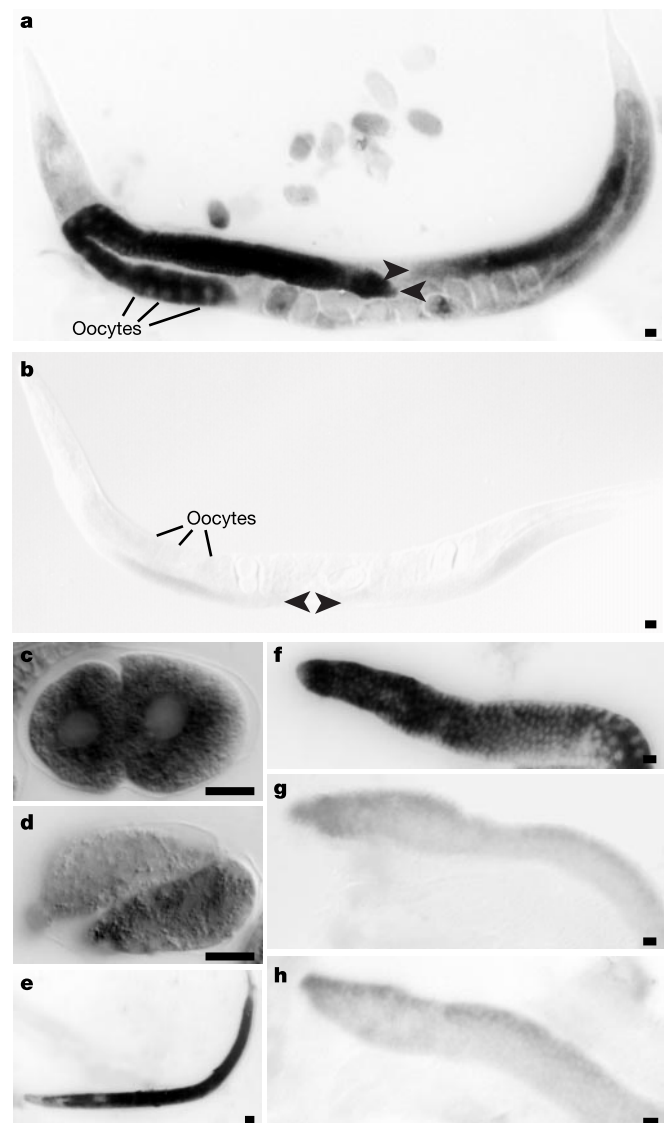


Figure 1 *cul-4* mRNA levels. **a, b**, *In situ* hybridization of wild-type adult hermaphrodites with *cul-4* antisense (**a**) or control sense (**b**) probes. Arrowheads mark the distal ends of the gonad arms; anterior oocytes are labelled. **c–e**, *In situ* hybridization with antisense *cul-4* probe of a two-cell-stage embryo (**c**), twofold-stage embryo (**d**) and L2 larva (**e**). **f–h**, *In situ* hybridization of dissected gonads from wild-type adult hermaphrodites (**f, h**) or wild-type hermaphrodite injected with *cul-4* dsRNA (**g**) probed with antisense *cul-4* (**f, g**) or sense *cul-4* (**h**). Scale bars, 10 μ m.

from an inability of the somatic gonad lineage to produce distal tip cells (DTCs) that are required to prevent germ cells from entering meiosis⁹.

We measured the amount of genomic DNA in enlarged *cul-4* RNAi blast cells and observed markedly elevated DNA levels. Two to three days post-hatch, *cul-4* RNAi seam cells contain up to 100C DNA content, compared with the 2C of the wild type (Figs 2g, h and 3d). Three distinct mechanisms can generate increased ploidy (Fig. 3a): (1) failed mitosis, in which a failure of chromosome separation and cytokinesis causes cells to enter G1 phase with a doubled DNA content¹⁰; (2) endoreplication, in which cells bypass mitosis and enter the next cell cycle with doubled DNA content¹¹; and (3) re-replication due to origin re-firing, referred to here as 're-replication', in which cells remain in S phase and continuously re-initiate DNA replication¹.

These mechanisms can be distinguished by three criteria. First, the presence of mitotic entry distinguishes the failed mitosis mechanism from endoreplication and re-replication, neither of which entails mitosis (Fig. 3a). The percentage of mitotic cells was

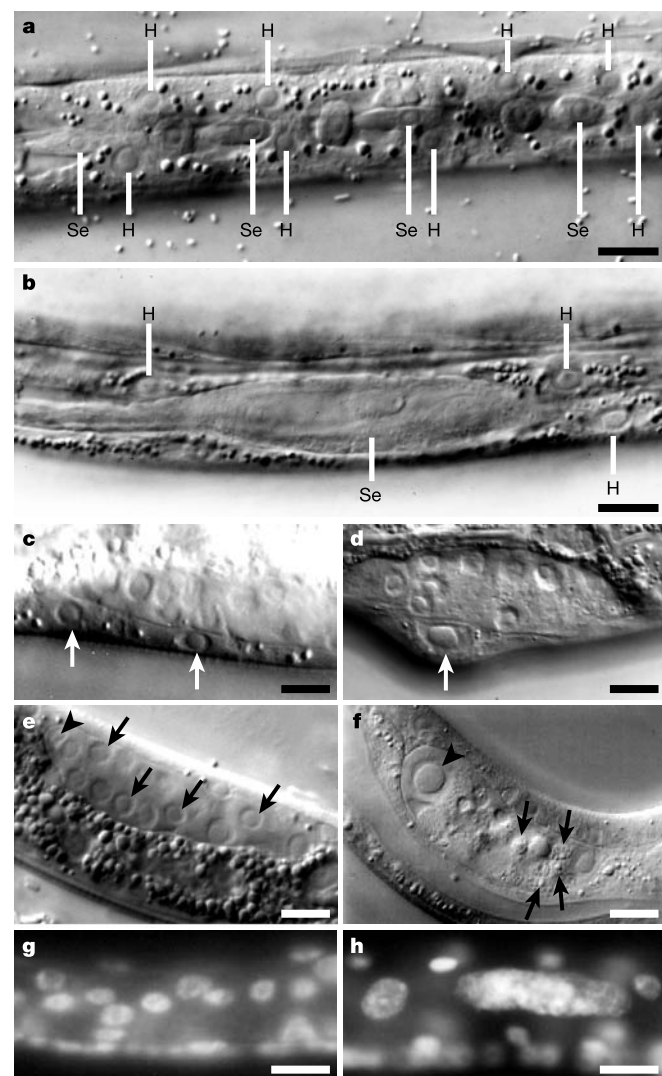


Figure 2 *cul-4* RNAi phenotype. **a, b**, DIC images of *clr-1(e1745)* (**a**) and *cul-4* RNAi, *clr-1(e1745)* (**b**) L2 larvae. Se, seam cell. H, hyp7 cell nuclei. **c, d**, DIC images of the ventral sides of wild-type (**c**) and *cul-4* RNAi (**d**) L2 larvae; arrows indicate P cells. **e, f**, Germ line of wild-type (**e**) and *cul-4* RNAi (**f**) L2 larvae. Arrowheads indicate DTC (**e**) or enlarged somatic gonad cell (**f**). Arrows indicate mitotic germ cells (**e**) or sperm (**f**). **g, h**, PI stain of genomic DNA in wild-type (**g**) and *cul-4* RNAi (**h**) L2 larvae. Scale bars, 10 μ m.

determined by immunodetection of the early mitotic marker phosphorylated histone H3 (ref. 12). In unsynchronized wild-type larvae, 3.8% of seam cells were mitotic (42/1116), whereas only 0.09% of *cul-4* RNAi seam cells were mitotic (1/1109), a 42-fold decrease (Fig. 3b). In addition, nuclear envelope breakdown (an early mitotic event) was not observed in the seam cell lineage of *cul-4* RNAi L1 larvae when followed by differential interference contrast (DIC) microscopy (0/3 larvae), but was readily apparent in wild-type L1 larvae (2/2). Therefore, the lack of mitotic entry precludes the failed mitosis mechanism.

Second, in both the failed mitosis and endoreplication mechanisms, cells do not arrest in S phase, but instead continue to cycle through G2 and G1 phases (Fig. 3a). The duration of S phase in

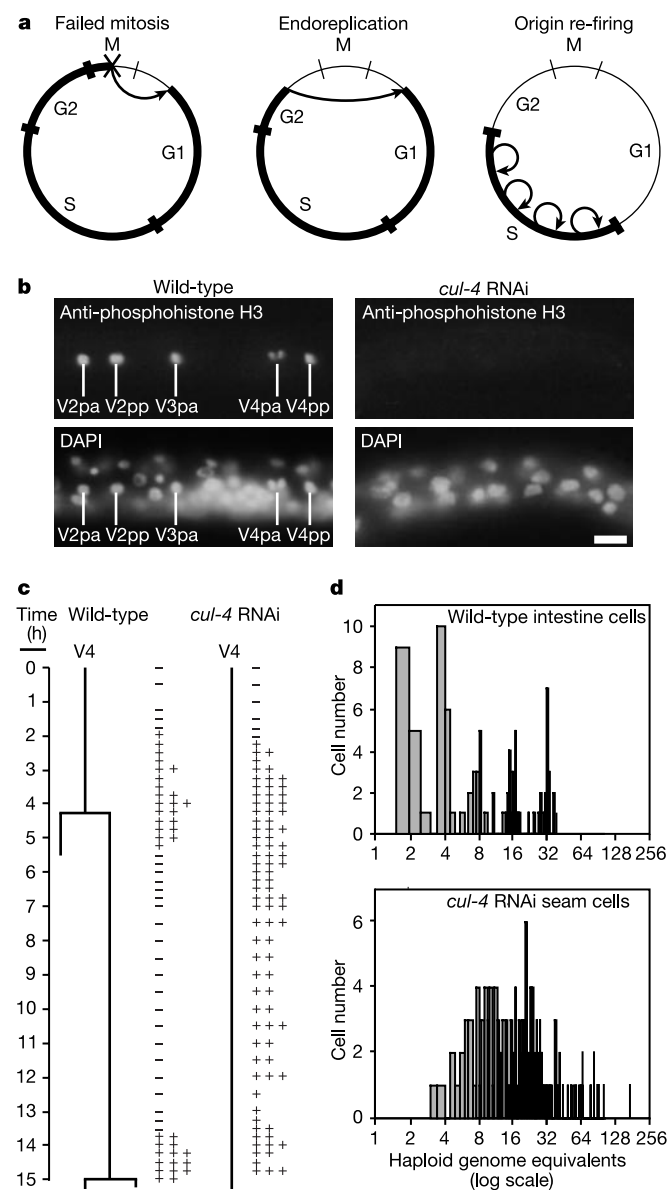


Figure 3 DNA re-replication in *cul-4* RNAi seam cells. **a**, Three mechanisms that generate increased ploidy. Dark lines represent the cell-cycle pathways. **b**, Anti-phosphorylated histone H3 and DAPI stain of wild-type and *cul-4* RNAi L2 larvae. Wild-type seam cells are named. Scale bar, 10 μ m. **c**, L1 lineage and *rnr::GFP* expression pattern of the V4 seam cell in wild-type and *cul-4* RNAi larvae. Minus symbols denote no *GFP* fluorescence. Plus symbols indicate the intensity of *rnr::GFP* fluorescence. **d**, Histogram of DNA content of intestine cells from unsynchronized wild-type larvae (top) and seam cells from unsynchronized *cul-4* RNAi larvae (bottom).

seam cells was measured using a reporter construct in which the S-phase-specific ribonucleotide reductase (*rnr*) promoter drives green fluorescent protein (GFP) expression¹³. We found that *rnr::GFP* expression began in both wild-type and *cul-4* RNAi seam cells at 2 h post-hatch (Fig. 3c). The wild-type *rnr::GFP* signal ended at approximately 5 h post-hatch, and did not resume until 14 h post-hatch, before the next cell division. By contrast, *cul-4* RNAi seam cells expressed *rnr::GFP* continuously through 15 h post-hatch (Fig. 3c). To further confirm that *cul-4* RNAi seam cells were in S phase during the period between wild-type mitotic divisions, we tested for incorporation of the thymidine analogue 5-bromodeoxyuridine (BrdU) into genomic DNA during a window of 5–11 h post-hatch. Wild-type seam cells did not incorporate BrdU during this time period (0/50); by contrast, 87.5% of *cul-4* RNAi seam cells incorporated BrdU (42/48). These results indicate that *cul-4* RNAi cells undergo an S-phase arrest, thereby supporting the re-replication model and arguing against the other two models.

Finally, in the failed mitosis and endoreplication mechanisms, increases in ploidy occur through doublings of genomic DNA, whereas in the re-replication mechanism ploidy increases are not quantized (Fig. 3a). The DNA content of *cul-4* RNAi seam cells was measured and found to increase continuously, rather than in doublings of 2C, as occurs in wild-type intestine cells that undergo endoreplication at each larval moult¹⁴ (Fig. 3d). Taken together,

these data suggest that the markedly increased ploidy in *cul-4* RNAi cells results from DNA re-replication.

The re-replication phenotype suggested a defect in DNA-replication licensing, which limits the extent of DNA replication in S phase¹. Cdt1 is a key replication-licensing factor in both yeast and metazoa^{2,3,15,16}. We identified a single *C. elegans* Cdt1 orthologue, *cdt-1*. In pairwise alignments, CDT-1 has 23%, 20%, 19% and 17% identity with its *Homo sapiens*, *Drosophila melanogaster*, *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae* orthologues^{2,15,17,18}, respectively. CDT-1 is required for DNA replication in *C. elegans*. *cdt-1* RNAi embryos arrest with approximately 60 cells and contain only trace amounts of DNA, indicating a virtually complete cessation of DNA replication, although early cell divisions occur (Fig. 4a). Unequal DNA segregation and DNA bridges between dividing cells are observed during these early cell divisions, presumably secondary consequences of the defective DNA replication (Fig. 4b).

We generated affinity-purified anti-CDT-1 antibody, which recognizes a single band of approximately the predicted size of 84.5 kDa on a western blot (Fig. 4c). Immunofluorescence with this antibody was used to determine the CDT-1 expression pattern. In adults, CDT-1 is present in germ cell nuclei and is enriched in oocyte nuclei (data not shown). In early embryos, CDT-1 is localized to chromosomes in mitotic cells during anaphase and early telophase (Fig. 4b; data not shown). CDT-1 nuclear staining is not present in S phase, which directly follows mitosis in the early embryo¹⁹ (Fig. 4b).

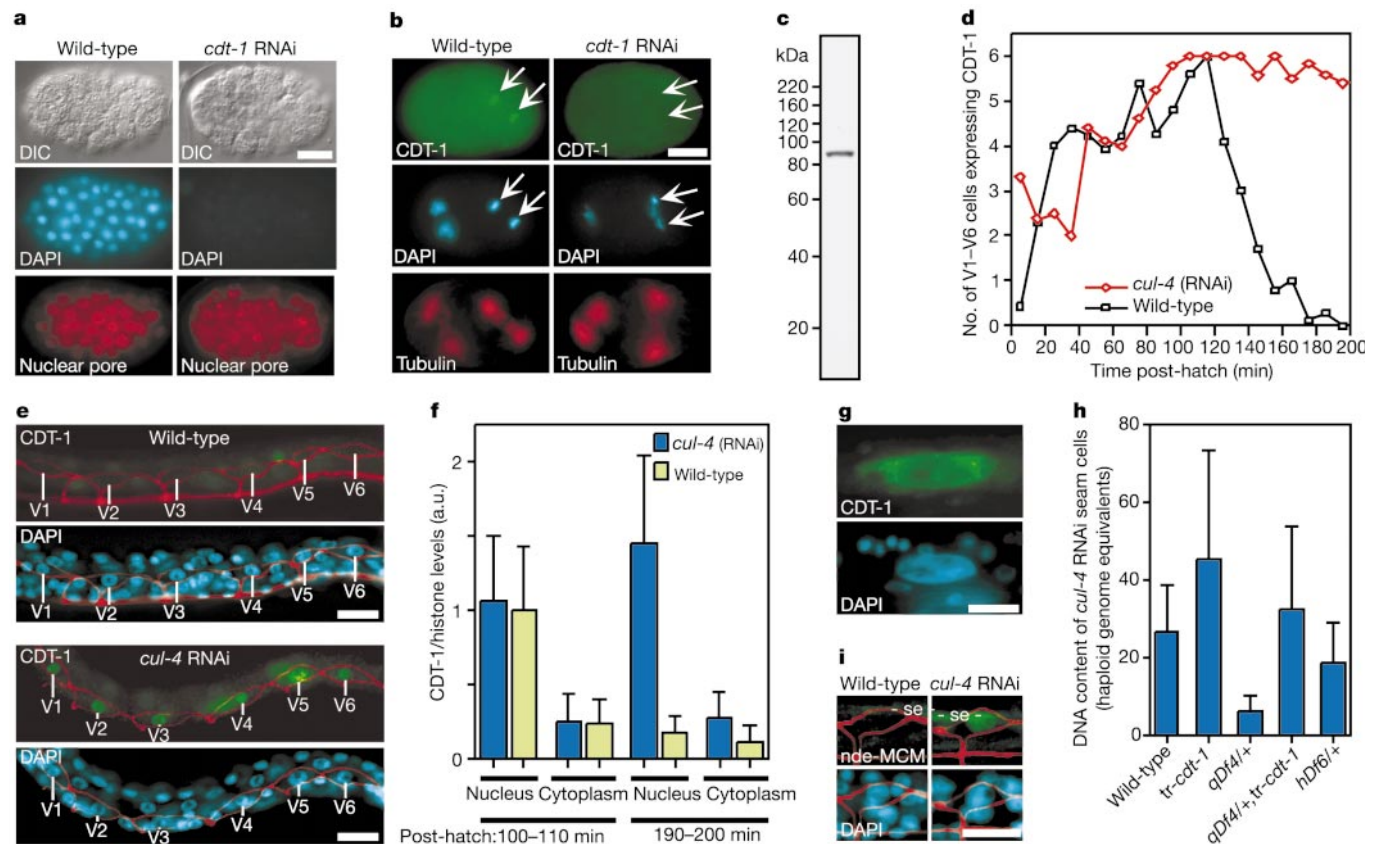


Figure 4 CDT-1 expression in *cul-4* RNAi animals. **a**, DIC, DAPI and anti-nuclear pore stain of 60-cell-stage wild-type and arrested *cdt-1* RNAi embryos. **b**, Anti-CDT1, DAPI and anti- α -tubulin stain of two-cell stage wild-type and *cdt-1* RNAi embryos. The *cdt-1* RNAi DAPI exposure is twice that of the wild type. Arrows indicate anaphase chromosomes. **c**, Anti-CDT-1 western blot of wild-type lysate. **d**, The average number of V1–V6 seam cells with nuclear CDT-1 in wild-type (squares) and *cul-4* RNAi (diamonds) animals at times post-hatch. **e**, Wild-type and *cul-4* RNAi larvae at 200 min post-hatch, stained with anti-CDT-1, DAPI and anti-AJM-1 (ref. 28) (red overlay, highlighting seam

cell boundaries). **f**, Anti-CDT-1 levels in V1–V6 seam cell nuclei and cytoplasm relative to anti-histone (a permeabilization control) in wild-type (green) and *cul-4* RNAi (blue) larvae at the times indicated. **g**, Anti-CDT-1 and DAPI stain of enlarged *cul-4* RNAi seam cell. **h**, DNA content of seam cells upon *cul-4* RNAi in the strains listed ('tr-*cdt-1*' denotes a strain expressing transgenic *cdt-1*). **i**, Non-detergent-extractable anti-MCM and DAPI stain of wild-type and *cul-4* RNAi seam cells at 225 min post-hatch, with anti-AJM-1 overlaid (red). Non-specific anti-MCM staining of the hypodermal seam ridge is marked (-se). Scale bars, 10 μ m.

Anti-CDT-1 staining in embryos is abolished by *cdt-1* RNAi, indicating the validity of the staining pattern (Fig. 4b).

Newly hatched larvae lack detectable CDT-1 expression. However, CDT-1 levels are transiently increased in the nuclei of blast cells before their mitotic divisions. We examined the timing of CDT-1 expression in the first cell divisions of the V1–V6 seam cells, which enter S phase at 2 h post-hatch (see Fig. 3c for timing). In the wild type, CDT-1 expression began approximately 20 min after hatching in a subset of seam cells, and by 120 min post-hatch all seam cells expressed CDT-1 at high levels (Fig. 4d, f). At 130 min post-hatch, the number of seam cells with CDT-1 protein dropped precipitously, and by 150 min post-hatch most seam cells no longer had CDT-1 protein (Fig. 4d–f). Significantly, the decrease in CDT-1 levels in S-phase cells occurred in both the nucleus and cytoplasm ($P < 1 \times 10^{-10}$ and $P < 0.0005$, respectively), indicating that CDT-1 does not undergo subcellular redistribution, but instead is degraded (Fig. 4f). A similar rapid disappearance of Cdt1 in S-phase cells is observed in fission yeast and mammalian cells, and in the latter the degradation of Cdt1 has been shown to occur through the ubiquitin proteolytic pathway; however, the ubiquitin ligase involved remains unknown^{2,20}.

In *cul-4* RNAi animals, there was faint CDT-1 expression in some seam cells at hatch, but this expression increased with a time course similar to that in wild-type animals, so that at 120 min post-hatch all seam cells expressed CDT-1 with a staining intensity similar to that of wild-type cells (Fig. 4d, f). However, in marked contrast to wild type, CDT-1 levels did not drop after 120 min post-hatch, remaining at elevated levels through the S phase (Fig. 4d–f). In 2-day-old arrested *cul-4* RNAi larvae, approximately half of the enlarged cells were actively synthesizing DNA, as shown by the incorporation of BrdU, and of these cells, 98% (97/99) had CDT-1 protein, further demonstrating that CDT-1 is present in S-phase *cul-4* RNAi cells. The CDT-1 protein in re-replicating cells is generally nuclear, but can also be present in both the nucleus and cytoplasm (Fig. 4g). These results indicate that CUL-4 is required for the rapid decrease of CDT-1 levels in S-phase cells.

If a failure to eliminate CDT-1 from S-phase cells were contributing to the *cul-4* RNAi re-replication phenotype, then overexpression of CDT-1 would be expected to enhance the phenotype, and conversely, removal of one copy of the *cdt-1* gene might reduce the level of CDT-1 sufficiently to suppress the phenotype. Attempts to overexpress *cdt-1* using microinjection to introduce transgenic *cdt-1* expressed from its own promoter or the heatshock *hsp16-2* promoter failed due to lethality among the F1 progeny (data not shown). Subsequently, microparticle bombardment, which produces chromosomal integrations of single or few copies²¹, was used to create a viable strain carrying transgenic *cdt-1* expressed from its own promoter. While *cul-4* RNAi produced seam cells with an average of 27C DNA content in wild-type larvae 2 days post-hatch, *cul-4* RNAi in the *cdt-1* overexpression strain produced an average of 45C DNA content, indicating a significant enhancement of the re-replication phenotype ($P < 0.05$) (Fig. 4h).

We tested for suppression of the *cul-4* RNAi phenotype by analysing the extent of re-replication in a strain heterozygous for the *qDf4* deficiency, which removes the *cdt-1* gene. Strikingly, the *cul-4* RNAi re-replication phenotype was suppressed fivefold in this strain ($P < 1 \times 10^{-6}$) (Fig. 4h). Reintroduction of the *cdt-1* gene into the *qDf4* heterozygous strain abolished the suppression, indicating that this was due to the loss of one copy of the *cdt-1* gene (Fig. 4h). By contrast, there was no significant suppression of the *cul-4* RNAi re-replication phenotype in a strain heterozygous for the *hDf6* deficiency, which deletes the orthologue of the replication-licensing factor Cdc6 (Fig. 4h).

In yeast and vertebrates, Cdt1 functions to license DNA for replication by promoting the binding of the Mcm2-7 putative replication helicase complex to replication origins to form the pre-replicative complex¹. In vertebrates, Mcm2-7 remains localized

to the nucleus after DNA replication, but dissociates from chromatin and can be extracted from nuclei with detergent²². Therefore, the presence of non-detergent extractable Mcm2-7 (nde-MCM) is an indication that chromatin has been licensed for replication. In early *C. elegans* embryos, all nuclei stain positive with an antibody that recognizes a conserved region of the MCM2-7 proteins. However, on detergent extraction, anti-MCM2-7 staining is only observed in telophase nuclei and smaller S-phase nuclei, consistent with MCM2-7 removal from replicated chromatin (data not shown). In larvae, nde-MCM is present in seam cell nuclei with a time course that is slightly delayed relative to nuclear CDT-1. At 120–135 min post-hatch, 95% of wild-type and 92% of *cul-4* RNAi seam cells had nde-MCM ($n = 84$ and 78 , respectively). Wild-type seam cells lost nde-MCM after S phase, with only 6% positive for nde-MCM at 225–240 min post-hatch ($n = 102$). By contrast, 81% of *cul-4* RNAi seam cells had nde-MCM at 225–240 min post-hatch ($n = 120$) (Fig. 4i). These results indicate that MCM2-7 remains associated with chromatin in re-replicating *cul-4* RNAi seam cells. Taken together, our results suggest that CUL-4 maintains genome stability by removing CDT-1 from S-phase cells, which ensures that CDT-1 is not available to reload MCM2-7 onto chromatin to re-initiate replication origin firing.

In the early embryonic cell divisions of *Xenopus*, Cdt1 is regulated by association with the inhibitory protein Geminin^{18,23}. Our observations indicate that the degradation of CDT-1 during S phase is crucial for the prevention of DNA re-replication, and they suggest that other CDT-1 control pathways, if present, cannot compensate for a failure of this central regulatory mechanism. In other eukaryotes, redundant controls operate to restrain DNA-replication licensing¹. In yeast, disruption of multiple controls is required to produce substantial re-replication^{24,25}. That inactivation of only a single gene, *cul-4*, leads to massive re-replication implies that CUL-4 is likely to regulate multiple aspects of DNA-replication licensing. □

Methods

Genetics and RNAi

The following strains of *C. elegans* were used: N2, wild type; CB3241 (*clr-1(e1745)*), VT765 (*unc-36(e251)*, *mal-103(unc-36(+))*, *rnr::GFP*); KR926 (*hDf6*, *dpy-5(e61)*, *unc-13(e450)/szT1*, *unc-3(e151)/szT1*); JR667 (*unc-119(e2498::Tc1)*, *wls51(unc-119(+))*, seam cell GFP marker); ET117 (*unc-119(ed3)*, *ekIs1(pID2.02/cdt-1)*); ET121 (*qDf4/unc-11(e47)*); and ET122 (*qDf4/unc-11(e47)*, *ekIs1(pID2.02/cdt-1)*). The following strains were used in Fig. 4h: N2; ET117; ET121; ET122; and KR926. The strains heterozygous for deficiencies *qDf4* and *hDf6* appear to be wild type, with normal DNA levels (data not shown). Plasmid pID2.02/*cdt-1* was created by cloning genomic *cdt-1*, including 727 base pairs upstream of the translational start (Y54E10A sequences 69,104–63,789, GenBank accession number AC024810), into plasmid pID2.02, which contains genomic *unc-119* sequence. pID2.02/*cdt-1* was introduced into *unc-119(ed3)* homozygotes by microparticle bombardment as described²¹. RNAi was performed using double-stranded RNA (dsRNA) derived from complementary DNA clones for *cul-4* (yk34c8, GenBank accession number D36543) and *cdt-1/Y54E10A.15* (yk10c5, GenBank accession number D34651). RNA was synthesized with Ambion T3 and T7 MegaScript kits. dsRNA was annealed and injected at a concentration of 0.5–1 mg ml⁻¹ into L4 larvae or young adults as described⁶, and progeny were analysed. Similar phenotypes were also obtained with the RNAi feeding method²⁶.

Antibodies and immunodetection

The *C. elegans cdt-1* and *cdc-6* genes were identified using BLAST²⁷ searches of National Center for Biotechnology Information databases for homologues of yeast and human Cdt1 and Cdc6. Anti-CDT-1 sera were produced in rabbits using purified recombinant His-tagged full-length CDT-1 fusion protein expressed from the pET15b vector (Novagen). Affinity purification was performed with His-CDT-1 fusion protein linked to activated CL4B Fast flow Sepharose (Amersham). Affinity-purified anti-CDT-1 antibodies from two different rabbits gave the same pattern of immunofluorescence staining. Other antibodies used were directed against: α -tubulin (N356, Amersham); nuclear pore (Mab414, BabCo); the adhesion junction component AJM-1 (ref. 28) (MH27, Developmental Studies Hybridoma Bank); BrdU (Sigma); phosphorylated histone H3 (Upstate Biotechnology); histone (Chemicon International); and pan-MCM (BD Pharmingen). Secondary antibodies were anti-rabbit and anti-mouse Alexa Fluor 488, 546 and 633 (Molecular Probes). Western blot with anti-CDT-1 antibody was performed with mixed-stage wild-type lysate separated on a NuPAGE 4–12% acrylamide gel (Invitrogen) and visualized by chemiluminescence using the SuperSignal West Femto substrate (Pierce). Immunofluorescence was performed on animals fixed using the 'freeze-crack' method as described²⁹. DNA was stained with either 1 μ g ml⁻¹ 4,6-diamidino-2-

phenylindole (DAPI) or with 50 $\mu\text{g ml}^{-1}$ propidium iodide (PI) after 20 $\mu\text{g ml}^{-1}$ RNase A treatment for 1 h at 37 °C. For analysis of CDT-1 expression in larvae at set times after hatching, pretzel-stage embryos were collected, and at 10-min intervals hatched larvae were transferred to plates with OP50 bacteria. An average of 9.4 larvae (range, 5–27) were analysed for each time point. In the anti-CDT-1 images of Fig. 4e, out-of-focus light near the seam cells derives from CDT-1-positive intestine cells located below the seam cells. For detergent extraction of MCM proteins, slides were freeze-fractured, incubated for 5 min in a 4 °C solution of 10 mM HEPES-KOH, pH 7.4, 300 mM sucrose, 100 mM NaCl, 3 mM MgCl_2 , 0.5% Triton X-100 and protease inhibitor cocktail (Roche), then processed for immunofluorescence according to the freeze-crack protocol²⁹.

In situ hybridization

Antisense and sense digoxigenin-labelled RNA probes were created from full-length *cul-4* cDNA using Ambion MegaScript T7 and T3 kits with digoxigenin-11-UTP (Roche). *In situ* hybridization was performed on whole animals, embryos or gonad arms dissected from hermaphrodites 2 days post-injection as described³⁰. Quantification of *cul-4* *in situ* hybridization levels in dissected gonads revealed: 100 $\pm$ 23 arbitrary units (a.u.) for wild-type antisense; 11 $\pm$ 28 a.u. for *cul-4* RNAi antisense; and 0 $\pm$ 16 a.u. for wild-type sense, n = 10 for each.

Microscopy

Animals were observed by DIC and immunofluorescence microscopy using a Zeiss Axioskop microscope. Images were taken with either TechPan film (Kodak) or a Hamamatsu ORCA-ER digital camera with Openlab 3.0.8 software (Improvision). Images were processed with Adobe Photoshop 6.0. Matched images were taken with the same exposure and processed identically. Matched images of anti-CDT-1, anti-AJM-1 and DAPI for Fig. 4e were deconvolved to equivalent extents using multineighbour deconvolution (Openlab). For quantification of GFP expression, generally two to three animals were observed per time point and the signals were averaged. Note that *rnr::GFP* signal persists beyond S phase in wild-type cells because of the perdurance of GFP protein. DNA was quantified from serial confocal images of PI-stained cells as described³⁰. For the seam cell DNA quantification in Fig. 4h, n = 15 for each strain. Anti-CDT-1 and anti-histone staining (Fig. 4f) were quantified from confocal images; n = 33, 19, 8 and 39 for wild-type 100–110 min, wild-type 190–200 min, *cul-4* RNAi 100–110 min, and *cul-4* RNAi 190–200 min post-hatch, respectively. Means are presented with standard deviations. Statistical significance was determined with Student's *t*-test (two-tailed, equal variance).

Received 5 February; accepted 8 April 2003; doi:10.1038/nature01747.

- Blow, J. J. & Hodgson, B. Replication licensing—defining the proliferative state? *Trends Cell Biol.* **12**, 72–78 (2002).
- Nishitani, H., Lygerou, Z., Nishimoto, T. & Nurse, P. The Cdt1 protein is required to license DNA for replication in fission yeast. *Nature* **404**, 625–628 (2000).
- Maio, R. D., Moreau, J. & Mechali, M. XCDT1 is required for the assembly of pre-replicative complexes in *Xenopus laevis*. *Nature* **404**, 622–625 (2000).
- Nayak, S. *et al.* The *Caenorhabditis elegans* Skp1-related gene family: diverse functions in cell proliferation, morphogenesis, and meiosis. *Curr. Biol.* **12**, 277–287 (2002).
- Tyters, M. & Jorgensen, P. Proteolysis and the cell cycle: with this RING I do thee destroy. *Curr. Opin. Genet. Dev.* **10**, 54–64 (2000).
- Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
- Hedgecock, E. M., Culotti, J. G. & Hall, D. H. The *unc-5*, *unc-6*, and *unc-40* genes guide circumferential migrations of pioneer axons and mesodermal cells on the epidermis in *C. elegans*. *Neuron* **4**, 61–85 (1990).
- Sulston, J. E. & Horvitz, H. R. Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Dev. Biol.* **56**, 110–156 (1977).
- Seydoux, G. & Schedl, T. The germline in *C. elegans*: origins, proliferation, and silencing. *Int. Rev. Cytol.* **203**, 139–185 (2001).
- Lorson, M. A., Horvitz, H. R. & van den Heuvel, S. LIN-5 is a novel component of the spindle apparatus required for chromosome segregation and cleavage plane specification in *Caenorhabditis elegans*. *J. Cell Biol.* **148**, 73–86 (2000).
- Edgar, B. A. & Orr-Weaver, T. L. Endoreplication cell cycles: more for less. *Cell* **105**, 297–306 (2001).
- Hendzel, M. J. *et al.* Mitosis-specific phosphorylation of histone H3 initiates primarily within pericentromeric heterochromatin during G2 and spreads in an ordered fashion coincident with mitotic chromosome condensation. *Chromosoma* **106**, 348–360 (1997).
- Hong, Y., Roy, R. & Ambros, V. Developmental regulation of a cyclin-dependent kinase inhibitor controls postembryonic cell cycle progression in *Caenorhabditis elegans*. *Development* **125**, 3585–3597 (1998).
- Hedgecock, E. M. & White, J. G. Polyploid tissues in the nematode *Caenorhabditis elegans*. *Dev. Biol.* **107**, 128–133 (1985).
- Whittaker, A. J., Royzman, I. & Orr-Weaver, T. L. *Drosophila* double parked: a conserved, essential replication protein that colocalizes with the origin recognition complex and links DNA replication with mitosis and the down-regulation of S phase transcripts. *Genes Dev.* **14**, 1765–1776 (2000).
- Cimborra, D. M. & Groudine, M. The control of mammalian DNA replication: a brief history of space and timing. *Cell* **104**, 643–646 (2001).
- Tanaka, S. & Diffley, J. F. Interdependent nuclear accumulation of budding yeast Cdt1 and Mcm2-7 during G1 phase. *Nature Cell Biol.* **4**, 198–207 (2002).
- Wohlschlegel, J. A. *et al.* Inhibition of eukaryotic DNA replication by Geminin binding to Cdt1. *Science* **290**, 2309–2312 (2000).
- Edgar, L. G. & McGhee, J. D. DNA synthesis and the control of embryonic gene expression in *C. elegans*. *Cell* **53**, 589–599 (1988).
- Nishitani, H., Taraviras, S., Lygerou, Z. & Nishimoto, T. The human licensing factor for DNA replication Cdt1 accumulates in G1 and is destabilized after initiation of S-phase. *J. Biol. Chem.* **276**, 44905–44911 (2001).

- Praitis, V., Casey, E., Collar, D. & Austin, J. Creation of low-copy integrated transgenic lines in *Caenorhabditis elegans*. *Genetics* **157**, 1217–1226 (2001).
- Todorov, I. T., Attaran, A. & Kearsley, S. E. BM28, a human member of the MCM2-3-5 family, is displaced from chromatin during DNA replication. *J. Cell Biol.* **129**, 1433–1445 (1995).
- Tada, S., Li, A., Maiorano, D., Mechali, M. & Blow, J. J. Repression of origin assembly in metaphase depends on inhibition of RLF-B/Cdt1 by geminin. *Nature Cell Biol.* **3**, 107–113 (2001).
- Nguyen, V. Q., Co, C. & Li, J. J. Cyclin-dependent kinases prevent DNA re-replication through multiple mechanisms. *Nature* **411**, 1068–1073 (2001).
- Yanow, S. K., Lygerou, Z. & Nurse, P. Expression of Cdc18/Cdc6 and Cdt1 during G2 phase induces initiation of DNA replication. *EMBO J.* **20**, 4648–4656 (2001).
- Timmons, L. & Fire, A. Specific interference by ingested dsRNA. *Nature* **395**, 854 (1998).
- Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
- Koppen, M. *et al.* Cooperative regulation of AJM-1 controls junctional integrity in *Caenorhabditis elegans* epithelia. *Nature Cell Biol.* **3**, 983–991 (2001).
- Miller, D. M. & Shakes, D. C. In *Caenorhabditis elegans: Modern Biological Analysis of an Organism* (eds Epstein, H. F. & Shakes, D. C.) 365–394 (Academic Press, San Diego, 1995).
- Feng, H. *et al.* CUL-2 is required for the G1-to-S phase transition and mitotic chromosome condensation in *Caenorhabditis elegans*. *Nature Cell Biol.* **1**, 486–492 (1999).

Acknowledgements We thank Y. Kohara, A. Fire, G. Seydoux, R. Roy and V. Ambros for reagents; the *Caenorhabditis* Genetics Center for nematode strains; R. Santurri for technical help; and H. Cai and P. Shen for critical reading of the manuscript. This work was supported by a grant from the National Institutes of Health to E.T.K.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to E.T.K. (ekipreos@cb.uga.edu).

RecBCD enzyme is a DNA helicase with fast and slow motors of opposite polarity

Andrew F. Taylor & Gerald R. Smith

Fred Hutchinson Cancer Research Center, Seattle, Washington 98109-1024, USA

Helicases are molecular motors that move along and unwind double-stranded nucleic acids¹. RecBCD enzyme is a complex helicase and nuclease, essential for the major pathway of homologous recombination and DNA repair in *Escherichia coli*². It has sets of helicase motifs¹ in both RecB and RecD, two of its three subunits. This rapid, highly processive enzyme unwinds DNA in an unusual manner: the 5'-ended strand forms a long single-stranded tail, whereas the 3'-ended strand forms an ever-growing single-stranded loop and short single-stranded tail. Here we show by electron microscopy of individual molecules that RecD is a fast helicase acting on the 5'-ended strand and RecB is a slow helicase acting on the 3'-ended strand on which the single-stranded loop accumulates. Mutational inactivation of the helicase domain in RecB or in RecD, or removal of the RecD subunit, altered the rates of unwinding or the types of structure produced, or both. This dual-helicase mechanism explains how the looped recombination intermediates are generated and may serve as a general model for highly processive travelling machines with two active motors, such as other helicases and kinesins.

Most helicases form Y-shaped molecules, with two equal-length single-stranded tails, during the unwinding of duplex DNA, but RecBCD enzyme forms structures with a single-stranded loop and two single-stranded tails. This unusual topology was demonstrated by electron microscopy of reaction intermediates³, as shown in Fig. 1a. Both the extent of unwinding and the size of the loop increase linearly with time for tens of kilobases, at rates of ~370 and ~150 nucleotides s⁻¹, respectively³ (see also Fig. 3a). The ever-

increasing lengths of both the loop and the short tail during unwinding suggest that the passage of single-stranded DNA from the loop, through the enzyme, to the short tail is a directional, and hence active (ATP-dependent), process.

It has been proposed that some helicases act through two separate activities, the melting of base pairs and the movement of the enzyme along one strand of the resultant single-stranded DNA⁴. Alternating use of these two activities both unwinds the DNA and moves the enzyme along it. In this view, the unwinding of duplex DNA by RecBCD enzyme would be mediated by a fast helicase, whereas the strand movement needed to move single-stranded DNA from the loop to the short tail would be provided by the motor component of a second, slower helicase. The RecB and RecD subunits each have canonical helicase domains and belong to superfamily 1 (SF1) of helicases¹. Isolated RecB and RecD subunits each have helicase activity and DNA-dependent ATPase activity^{5–7}. Mutational alterations (RecB^{K29Q} and RecD^{K177Q}) in the Walker A box of the helicase motifs¹ of RecB and RecD decrease ATPase activities of individual

subunits or reconstituted holoenzyme at least 50-fold^{6–8}. Alteration of the lysine (Lys 29 or Lys 177) in this motif, which is invariant in helicases and many other ATP-binding proteins⁹, inactivates the ATPases without affecting the helicases' ability to bind DNA¹⁰, suggesting that this alteration specifically inactivates the motor activity. Heterotrimeric RecBCD enzyme, with one copy of each polypeptide, can unwind DNA¹¹. Collectively, this information indicated that RecB and RecD might be two helicases that act in concert to make the loop-tail unwinding structures described previously³. Here we used derivatives of RecBCD enzyme, mutated in one or the other helicase activity or lacking one helicase subunit, to test the hypothesis that these unwinding structures result from the combined actions of a fast and a slow helicase motor. Unwinding of double-stranded DNA by RecB^{K29Q}CD enzyme, in which RecD is the only active ATPase, produced a novel unwinding intermediate, with a terminal single-stranded loop accompanied by only one single-stranded tail (Fig. 1b; Supplementary Figs S1 and S2). We term this a 'loop one-tail' structure, to distinguish it from the wild-type 'loop two-tail' structure (Fig. 1a), previously referred to as a 'loop-tails' structure³. The loop and the tail of the loop one-tail structures had equivalent lengths (Fig. 1c), indicating that these structures did not result from nucleolytic removal of a second tail.

We compared the strand polarity of the unwinding structures made by RecB^{K29Q}CD enzyme with that of structures made by the wild-type enzyme, using DNA with biotinyl moieties at the 3' or the 5' end and anti-biotin antibody conjugated to gold particles. In structures made by the wild-type enzyme the long tail had a 5' end (Fig. 1a; Supplementary Fig. S3), as reported previously¹²; similar loop two-tail structures with the same strand polarity were made under a variety of reaction conditions by using 3' end-labelled substrates (Supplementary Fig. S4). In structures made by the RecB^{K29Q}CD mutant enzyme the single tail had a 5' end (Fig. 1b and Supplementary Fig. S2), showing the absence of 5' end degradation as inferred from the length measurements noted above. The nearly wild-type rate of unwinding by RecB^{K29Q}CD enzyme (Fig. 1d; see also Fig. 3a) shows that the RecD helicase can by itself provide the motive power to rapidly move the enzyme and unwind DNA. We infer that the inactive RecB^{K29Q} subunit remained bound to the 3' end and that it initially bound, producing a loop but no second tail.

We addressed the contribution of the RecB helicase motor to the unwinding activity in two ways, with RecBC enzyme lacking the RecD subunit but retaining helicase activity¹³, and with RecBCD^{K177Q} enzyme lacking the RecD ATPase⁶. The reaction of double-stranded DNA with RecBC enzyme produced only Y-shaped molecules with two equal-length single-stranded tails (~80%; Fig. 2a) or molecules that could be derived from them by mechanical shear (~20%). Because RecC protein lacks the set of helicase motifs¹ and has no detectable ATPase activity¹⁴, we infer that the single RecB helicase of RecBC enzyme unwinds DNA but without loop formation. In an assay that couples a single-strand-specific exonuclease to the helicase activity, RecBC unwinds DNA only about 20% as rapidly as RecBCD¹³; our electron microscopy measurements agree with this estimate (data not shown). Thus, RecD is necessary for fast unwinding; RecB, complexed with RecC, is a slower helicase.

The contribution of the RecB motor was also addressed by using RecBCD^{K177Q} enzyme, which lacks an active RecD motor. This mutant enzyme unwound DNA at ~73 nucleotides s⁻¹ (Fig. 3a), about 20% of the rate of the wild-type enzyme (~370 nucleotides s⁻¹); these relative rates of the wild-type and RecBCD^{K177Q} enzymes are similar to those found with the exonuclease-coupled assay¹³ and by gel-electrophoresis assays of unwinding (data not shown). Thus, either removal of RecD or inactivation of its ATPase resulted in an unwinding rate about 20% of that of the wild-type enzyme, which is consistent with the conclusion above that RecB is a slower helicase than RecD.

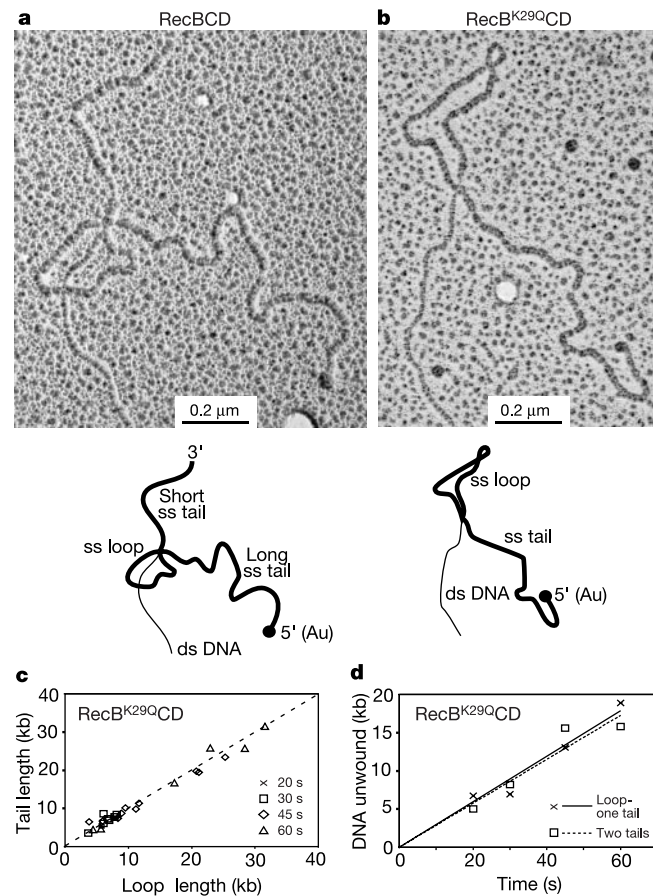


Figure 1 RecB^{K29Q}CD enzyme, with an inactive RecB helicase, unwinds DNA by means of a loop and a 5'-terminated tail. **a, b**, Electron micrographs of unwinding intermediates, produced by reaction of 0.7 nM 5'-biotinylated double-stranded DNA with 4 nM RecBCD enzyme or RecB^{K29Q}CD enzyme for 60 s. As identified below the micrographs, the thin sinuous lines are double-stranded DNA, the thicker ones are SSB-coated single-stranded DNA, and the electron-dense particles are anti-biotin antibody-conjugated gold beads that identify the 5' termini of the unwound DNA. **a**, A wild-type loop two-tail unwinding intermediate produced by incubation with RecBCD enzyme. **b**, A loop one-tail unwinding intermediate produced by incubation with RecB^{K29Q}CD enzyme. ds, double-stranded; ss, single-stranded. **c**, The single-stranded loop and the tail made by RecB^{K29Q}CD enzyme are of equal lengths. All the data used in **d** are plotted. The dashed line depicts equality. kb, kilobases. **d**, Travel rate of loop one-tail and two-tail structures produced by RecB^{K29Q}CD enzyme: rates are 300 nucleotides s⁻¹ (loop one-tail) and 290 nucleotides s⁻¹ (two-tails).

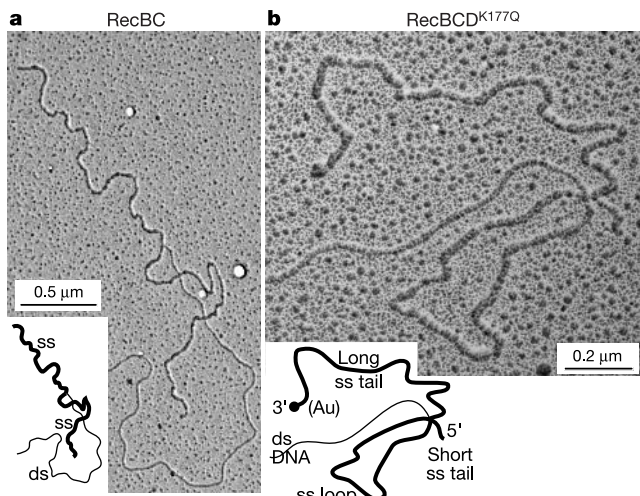


Figure 2 Enzymes without RecD or with mutant RecD (RecBCD^{K177Q}) unwind DNA slowly and with different topologies. **a**, Double-stranded λ DNA (0.2 nM) after reaction for 30 s with 80 nM RecBC enzyme, lacking the RecD subunit. No looped structures were observed among 80 partly unwound molecules. **b**, Loop two-tail structure produced by reaction of 1 nM 3'-biotinylated double-stranded DNA for 60 s with 2 nM RecBCD^{K177Q} enzyme: the gold particle on the 3' terminus shows that the strand polarity is opposite to that of the wild-type enzyme (Fig. 1a). ds, double-stranded; ss, single-stranded.

The strand polarity of the unwinding structures made by RecBCD^{K177Q} enzyme was opposite to that of the structures made by wild-type enzyme noted above (Fig. 1a). In 24 of 25 loop two-tail structures made by RecBCD^{K177Q} and bound to a gold particle, the longer tail had a 3' terminus, indicating that the loop and shorter tail were on the 5'-ended strand (Fig. 2b and Supplementary Fig. S3). The mean rate of growth of the 5' tail was 15 nucleotides s⁻¹ by RecBCD^{K177Q} and 370 nucleotides s⁻¹ by RecBCD (Fig. 3a). We infer that in the structures made by RecBCD^{K177Q} enzyme the inactive RecD subunit moves very slowly along the 5'-terminated strand, about 5% of its rate in the wild-type enzyme. This very slow movement of the RecD^{K177Q} subunit might result from residual, weak ATPase activity in that subunit or from force generated by the wild-type RecB subunit, perhaps by an allosteric conformational change of RecD upon ATP hydrolysis by RecB. Travel of the mutant RecD subunit requires an active RecB subunit, as the double-mutant enzyme RecB^{K29Q}CD^{K177Q} had no detectable unwinding or ATPase activity (Supplementary Table S1). Finally, we infer from the data in Fig. 3a that the active RecB motor

in RecBCD^{K177Q} enzyme unwinds the DNA, producing the longer single-stranded tail with a 3' end, at about the same rate (73 base pairs s⁻¹) that the RecB motor unwinds DNA in RecBC enzyme; this is about one-third of the rate at which we infer that RecB translocates along the 3'-ended single-stranded DNA strand in the RecBCD holoenzyme.

We examined the degree of coupling of the fast RecD and slow RecB motors in wild-type enzyme by comparing the distance travelled by the fast motor with that travelled by the slow motor in individual molecules (Fig. 3b). Whereas the mean lengths of both tails increased linearly with time (Fig. 3a), much variation in the relative rates of the two can be seen in individual molecules. There is therefore no tight coupling between the two motors, despite a constant rate of the two for the average behaviour of molecules. Figures 1c and 3b also illustrate an approximately twofold variation in unwinding rate of individual molecules, which has been seen by other techniques with the wild-type RecBCD enzyme^{15,16}.

How do the RecB and RecD helicases interact with DNA to produce the loop two-tail intermediates? In a simple model (Fig. 4), each of the two helicase motors travels along (translocates on) one or the other strand of the DNA; the faster helicase actively unwinds the DNA, and a single-stranded loop accumulates on the strand with the slower helicase. Other helicases of the SF1 and SF2 families, such as PcrA and PriA, can translocate along single-stranded DNA; that is, without actual unwinding^{17,18}. The strand polarities of the unwinding structures agree with the ultraviolet-dependent cross-linking of RecB to the 3'-ended strand and RecD to the 5'-ended strand in the RecBCD-DNA initiation complex¹⁹ (Fig. 4a) and with the inferred helicase polarities of isolated RecB and RecD^{5,7}. That RecD is a faster helicase than RecB is indicated by the slower unwinding of DNA by RecBCD^{K177Q} and RecBC enzymes than by wild-type RecBCD enzyme, and the similar fast rates by RecB^{K29Q}CD and RecBCD enzymes (Figs 1 and 3).

Thus, in this model the loop and short tail are predicted to be on the 3'-ended strand in DNA unwound by RecBCD enzyme, as observed (Fig. 1a, Supplementary Fig. S3 and ref. 12). Complete mutational inactivation of the RecB or RecD subunit would result in the mutant (inactive) subunit's remaining bound at or near its initial terminus, forming a loop on that strand and an equal-length single-stranded tail on the other strand (Fig. 4c, d), as was observed for RecB^{K29Q}CD (Fig. 1b). The very short 5'-terminated tails observed with RecBCD^{K177Q} (Fig. 2b) might result, as noted above, either from slight residual activity in the mutated ATPase or from interaction between active and inactive ATPases in that mutant. RecBC enzyme, having only one helicase, would make only forks (Fig. 4e), as observed (Fig. 2a). For RecBCD to form a loop two-tail structure (Figs 1a and 4b), the two motors (RecB and RecD)

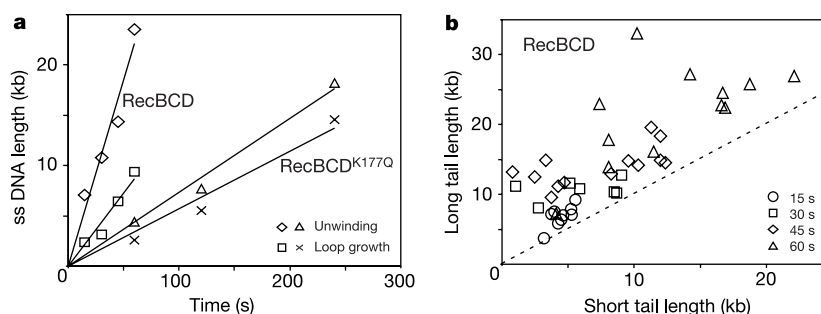


Figure 3 Analysis of DNA unwinding by RecBCD and RecBCD^{K177Q}. **a**, Slow unwinding and loop growth by RecBCD^{K177Q} enzyme. Mean unwinding and loop-growth rates are 370 and 150 nucleotides s⁻¹ for RecBCD enzyme and 73 and 57 nucleotides s⁻¹ for RecBCD^{K177Q} enzyme. ss, single-stranded. **b**, Lack of tight coupling between unwinding and single-stranded translocation by RecBCD enzyme. For each of the wild-type

molecules measured in **a**, the length of the long tail (representing unwinding) is plotted against that of the short tail (representing strand passage from loop to tail). The dashed line indicates equality; by definition of the short and long tails, the data points must be on or above this line.

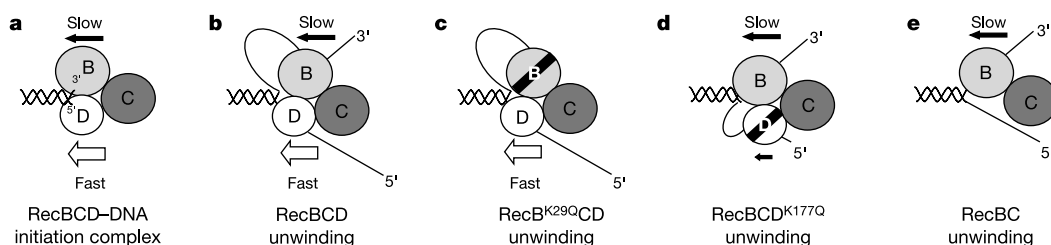


Figure 4 Model for unwinding by RecBCD enzyme. Arrows represent the active helicases in RecB and RecD; bars through circles depict mutant helicases (RecB^{K29Q}, RecD^{K177Q}) devoid of ATPase activity. **a**, Complex of RecBCD bound to a double-stranded DNA end in the absence of ATP. Irradiation with ultraviolet light crosslinks RecB to the 3'-ended strand, and RecC and RecD to the 5'-ended strand¹⁹, which is consistent with the inferred polarities of the RecB and RecD helicases^{5,7}. **b**, Wild-type holoenzyme travels at the speed

of RecD, the faster helicase; a loop of single-stranded DNA accumulates ahead of RecB, the slower helicase. **c**, The inactive RecB^{K29Q} subunit remains bound to the 3' end of the unwound strand, producing a loop one-tail intermediate. **d**, **e**, The lower velocity of the RecB helicase results in slow unwinding by RecBCD^{K177Q} (**d**) and RecBC (**e**) enzymes, the latter without any looped intermediates.

must be held together during unwinding. Although we know of no direct evidence for a direct RecB–RecD contact, the lack of nuclease activity in RecBC enzyme²⁰, which harbours the RecB nuclease domain but lacks RecD²¹, implies that RecB and RecD must interact to form the active nuclease, perhaps by direct contact.

Although the slower RecB helicase might contribute only modestly to the very high rate of unwinding catalysed by RecBCD enzyme, it is clearly crucial for many activities of the enzyme. *E. coli* cells harbouring RecB^{K29Q}CD enzyme are profoundly recombination deficient (conjugal recombination is decreased ~100-fold, as in *recB*-null or *recC*-null mutants), and modestly sensitive to ultraviolet radiation, whereas those harbouring RecBCD^{K177Q} enzyme are resistant to radiation and are only slightly impaired for conjugal recombination (S. K. Amundsen and G.R.S., unpublished observations). Similarly, RecBCD^{K177Q} enzyme is only slightly impaired in double-stranded and single-stranded DNA exonuclease²², but RecB^{K29Q}CD enzyme shows complete loss of double-stranded DNA exonuclease activity, although it retains single-stranded DNA exonuclease²³. A nuclease determinant is located within the carboxy-terminal domain of RecB²¹. The lack of double-stranded DNA exonuclease in RecB^{K29Q}CD enzyme is therefore consistent with the 3'-terminated tail, the strand more vigorously degraded by the wild-type enzyme²⁴, being translocated past that nuclease motif during its exit from the enzyme. However, the nuclease activity of RecBCD is not itself needed for recombination: cells bearing the (nuclease-deficient) RecBC enzyme are recombination proficient²⁰. The recombination deficiency of cells harbouring RecB^{K29Q}CD enzyme might therefore reflect the importance of the 3'-terminated single strand, uniquely lacking in this mutant, to homologous recombination. This is the strand on which the Chi hotspot sequence is recognized²⁵ and onto which RecA protein is loaded by RecBCD enzyme²⁶.

We have shown here that the two helicases of RecBCD enzyme, specified by the helicase domains in RecB and RecD, are both required for full activity of the enzyme. The helicase in RecD provides the fast unwinding activity. RecB, which has very weak helicase activity in isolation³, is a genuine but slow helicase when complexed with RecC alone (Figs 2a and 4e), but seems only to translocate single-stranded DNA from the loop to the 3'-terminated tail in RecBCD holoenzyme (Figs 1a and 4b). One model of helicase action⁴, with separate melting and translocation steps, fits well with this separation of RecB functions. RecBCD is highly processive: it can unwind more than 20 kilobases of DNA without dissociation^{3,16}. However, the isolated subunits have very low processivity, being able to unwind only about 50 base pairs^{5,7}. The exceptionally high processivity of RecBCD might stem at least in part from its having two motors: the probability that both would simultaneously dissociate from the DNA would be much less than that of one motor acting on its own. Other processive enzymes also have more than

one motor: some helicases have two or six identical subunits¹ and kinesins typically have two²⁷. Recent reports show that a monomeric helicase and a monomeric kinesin become active or more processive when converted to the dimeric form^{27,28}. Thus, RecBCD enzyme, like the human TFIIH factor involved in transcription and DNA repair²⁹, fits this model but has the unusual feature of having two non-identical motors with opposite DNA-strand polarities. This asymmetric feature might impart RecBCD enzyme's asymmetry in other aspects of its promotion of genetic recombination². □

Methods

Proteins and DNA

RecBCD enzyme was purified from a highly overexpressing strain³⁰ by chromatography on Amersham Biosciences Hi-Trap Heparin and BioRad Agarose A1.5m columns. RecB^{K29Q}CD and RecBCD^{K177Q} enzymes^{22,23} were prepared from strain V330 [Δ (*recC-argA*)234 λ^- F'] containing derivatives of plasmid pACYC184 bearing an 18.5-kilobase *Bam*HI fragment of *E. coli* carrying the mutant *recBCD* region (a gift from D. Julin). Enzymes were purified by chromatography on Amersham Biosciences HiTrap Q Sepharose, HiPrep Sephacryl S-300 HR and HiTrap Heparin columns. RecB and RecC were purified separately³⁰ with HiTrap Q Sepharose, BioRad CHT-II and HiTrap Heparin columns. RecBC enzyme was made by mixing RecB and RecC, with subsequent purification on a HiPrep Sephacryl S-300 HR column. *E. coli* single-strand DNA-binding protein (SSB) was from Promega. Bacteriophage λ was induced from lysogens, purified by two CsCl-gradient centrifugations, and the DNA (48 kilobases, with 12-nucleotide 5' overhangs) was extracted with phenol. Plasmid pDWS2 DNA (23 kilobases) was linearized with *Eco*RI and labelled at its 3' ends with bio-16-dUTP¹² (Roche), producing a blunt-ended double-stranded DNA substrate. DNA with 5'-biotin labels was produced by ligation of annealed oligonucleotides (5'-biotin-tetra-ethylene glycol-GGGCGCGACCTAGCT-3' and 5'-pAGGTCGCGCG-3'; IDT Inc.) to pDWS2 DNA linearized with *Sac*I, followed by sucrose-density-gradient sedimentation to remove unincorporated oligonucleotides. ϕ X174 RFII and viral DNA were from New England Biolabs.

Reactions

RecBCD, RecB^{K29Q}CD or RecBCD^{K177Q} enzyme was incubated with 20 mM MOPS-KOH pH 7.0, 1 mM MgCl₂, 1 mM CaCl₂, 1 mM dithiothreitol, 3 μ M SSB and DNA at 37 °C for 15 min, to allow binding. Reaction was started by adding ATP to 5 mM and stopped by adding EDTA to 40 mM; goat anti-biotin gold conjugate (10 nm; Ted Pella Inc.) was added where appropriate, and the reactions were fixed by incubation at 37 °C for 20 min with 0.2% glutaraldehyde. RecBC enzyme was incubated with 50 mM MOPS-KOH pH 7.0, 10 mM MgCl₂, 1 mM dithiothreitol, 3 μ M SSB and 0.2 nM λ DNA at 37 °C for 15 min, to allow binding. Reaction was started by adding ATP to 5 mM and stopped by adding EDTA to 100 mM. The products were diluted 1:4, fixed as above and dialysed overnight against 20 mM MOPS-KOH pH 7.0, 0.1 mM EDTA.

Microscopy and data analysis

Samples were spread for microscopy³ and examined in a JEOL 1010 transmission electron microscope. Images were acquired with a Gatan digital camera and DNA contour lengths measured by manual tracing using NIH Image software. Contour lengths were converted to nucleotides by reference to standard molecules, namely λ or pDWS2 double-stranded DNA and relaxed ϕ X174 double-stranded circles or SSB-coated ϕ X174 single-stranded circles. Loop one-tail or two-tail structures were identified as the junction of a double-stranded DNA segment with three or four SSB-coated single-stranded DNA arms, two of which were connected to form a loop. Y structures were identified as the junction of a double-stranded DNA segment and two linear SSB-coated single-stranded DNA arms. For each molecule the distance unwound was defined as the length of the longer continuous single strand: loop plus short tail versus long tail for RecBCD or RecBCD^{K177Q} enzymes, and loop versus tail for RecB^{K29Q}CD enzyme. Mean values for each time point were

plotted and rates were calculated by linear regression forced through the origin. Data points in Fig. 1c, d are from measurements of 2, 7, 10 and 6 molecules for the loop one-tail structures, and 17, 12, 12 and 10 molecules for the two-tail structures at 20, 30, 45 and 60 s, respectively, with 16 nM RecB^{K29Q}CD enzyme and 0.2 nM λ DNA. Data points in Fig. 3a are means of measurements of 9, 7, 12 and 13 molecules for the 15, 30, 45 and 60 s points with 0.5 nM wild-type enzyme and of 4, 3 and 6 molecules for the 60, 120 and 240 s points with 2 nM RecBCD^{K177Q} enzyme and 0.2 nM λ DNA. Typically, more than 90% of the observed partly unwound structures were interpretable. The topology of 651 of the 862 structures at double-stranded DNA ends seen with RecBCD was similar to the structure in Fig. 1a. The topology of 141 molecules, including 41 with gold at the end of the tail, of the 1,117 structures seen with RecB^{K29Q}CD was similar to the structure in Fig. 1b, which is unique to RecB^{K29Q}CD. The topology of 146 of the 258 structures seen with RecBCD^{K177Q} was similar to the structure in Fig. 2b. The remaining structures were predominantly forks (Fig. 2a), which could result either from release of the loop during unwinding or from failure of the glutaraldehyde fixation necessary to preserve partly unwound structures during preparation for microscopy.

Received 7 November 2002; accepted 14 April 2003; doi:10.1038/nature01674.

- Lohman, T. M. & Bjornson, K. P. Mechanisms of helicase-catalyzed DNA unwinding. *Annu. Rev. Biochem.* **65**, 169–214 (1996).
- Smith, G. R. Homologous recombination near and far from DNA breaks: Alternative roles and contrasting views. *Annu. Rev. Genet.* **35**, 243–274 (2001).
- Taylor, A. & Smith, G. R. Unwinding and rewinding of DNA by the RecBC enzyme. *Cell* **22**, 447–457 (1980).
- Velankar, S. S., Soultanas, P., Dillingham, M. S., Subramaya, H. S. & Wigley, D. B. Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. *Cell* **97**, 75–84 (1999).
- Phillips, R. J., Hicketon, D. C., Boehmer, P. E. & Emmerson, P. T. The RecB protein of *Escherichia coli* translocates along single-stranded DNA in the 3' to 5' direction: a proposed ratchet mechanism. *Mol. Gen. Genet.* **254**, 319–329 (1997).
- Chen, H.-W., Ruan, B., Yu, M., Wang, J.-d. & Julin, D. A. The RecD subunit of the RecBCD enzyme from *Escherichia coli* is a single-stranded DNA dependent ATPase. *J. Biol. Chem.* **272**, 10072–10079 (1997).
- Dillingham, M. S., Spies, M. & Kowalczykowski, S. C. RecBCD enzyme is a bipolar DNA helicase. *Nature* **423**, 893–897 (2003).
- Chen, H.-W., Randle, D. E., Gabbidon, M. & Julin, D. A. Functions of the ATP hydrolysis subunits (RecB and RecD) in the nuclease reactions catalyzed by the RecBCD enzyme from *Escherichia coli*. *J. Mol. Biol.* **278**, 89–104 (1998).
- Sarasante, M., Sibbold, P. R. & Wittinghofer, A. The P-loop—a common motif in ATP- and GTP-binding proteins. *Trends Biochem. Sci.* **15**, 430–434 (1990).
- George, J. W., Brosh, R. M. Jr & Matson, S. W. A dominant negative allele of the *Escherichia coli* *uvrD* gene encoding DNA helicase II. A biochemical and genetic characterization. *J. Mol. Biol.* **229**, 67–78 (1994).
- Taylor, A. F. & Smith, G. R. Monomeric RecBCD enzyme binds and unwinds DNA. *J. Biol. Chem.* **270**, 24451–24458 (1995).
- Braedt, G. & Smith, G. R. Strand specificity of DNA unwinding by RecBCD enzyme. *Proc. Natl Acad. Sci. USA* **86**, 871–875 (1989).
- Koranyi, F. & Julin, D. A. Efficiency of ATP hydrolysis and DNA unwinding by the RecBC enzyme from *Escherichia coli*. *Biochemistry* **33**, 9552–9560 (1994).
- Masterson, C. et al. Reconstitution of the activities of the RecBCD holoenzyme of *Escherichia coli* from the purified subunits. *J. Biol. Chem.* **267**, 13564–13572 (1992).
- Dohoney, K. M. & Gelles, J. χ -sequence recognition and DNA translocation by single RecBCD helicase/nuclease molecules. *Nature* **409**, 370–374 (2001).
- Bianco, P. R. et al. Processive translocation and DNA unwinding by individual RecBCD enzyme molecules. *Nature* **409**, 374–378 (2001).
- Dillingham, M. S., Wigley, D. B. & Webb, M. R. Direct measurements of single-stranded DNA translocation by PcrA helicase using the fluorescent base analogue 2-aminopurine. *Biochemistry* **41**, 643–651 (2002).
- Lee, M. S. & Marians, K. J. Differential ATP requirements distinguish the DNA translocation and DNA unwinding activities of the *Escherichia coli* PRI A protein. *J. Biol. Chem.* **265**, 17078–17083 (1990).
- Ganesan, S. & Smith, G. R. Strand-specific binding to duplex DNA ends by the subunits of *Escherichia coli* RecBCD enzyme. *J. Mol. Biol.* **229**, 67–78 (1992).
- Chaudhury, A. M. & Smith, G. R. A new class of *Escherichia coli* recBC mutants: Implications for the role of RecBC enzyme in homologous recombination. *Proc. Natl Acad. Sci. USA* **81**, 7850–7854 (1984).
- Yu, M., Souaya, J. & Julin, D. A. Identification of the nuclease active site in the multifunctional RecBCD enzyme by creation of a chimeric enzyme. *J. Mol. Biol.* **283**, 797–808 (1998).
- Koranyi, F. & Julin, D. A. Enzymatic effects of a lysine-to-glutamine mutation in the ATP-binding consensus sequence in the RecD subunit of the RecBCD enzyme from *Escherichia coli*. *J. Biol. Chem.* **267**, 1733–1740 (1992).
- Hsieh, S. & Julin, D. A. Alteration by site-directed mutagenesis of the conserved lysine residue in the consensus ATP-binding sequence of the RecB protein of *Escherichia coli*. *Nucleic Acids Res.* **20**, 5647–5653 (1992).
- Anderson, D. G. & Kowalczykowski, S. C. The recombination hot spot χ is a regulatory element that switches the polarity of DNA degradation by the RecBCD enzyme. *Genes Dev.* **11**, 571–581 (1997).
- Bianco, P. R. & Kowalczykowski, S. C. The recombination hot spot χ is recognized by the translocating RecBCD enzyme as the single strand of DNA containing the sequence 5'-GCTGTGG-3'. *Proc. Natl Acad. Sci. USA* **94**, 6706–6711 (1997).
- Anderson, D. G. & Kowalczykowski, S. C. The translocating RecBCD enzyme stimulates recombination by directing RecA protein onto ssDNA in a χ -regulated manner. *Cell* **90**, 77–86 (1997).
- Tomishige, M., Klopfeinstein, D. R. & Vale, R. D. Conversion of Unc104/KIF1A kinesin into a processive motor after dimerization. *Science* **297**, 2263–2267 (2002).
- Ha, T. et al. Initiation and re-initiation of DNA unwinding by the *Escherichia coli* Rep helicase. *Nature* **419**, 638–641 (2002).

- Berneburg, M. & Lehmann, A. R. Xeroderma pigmentosum and related disorders: defects in DNA repair and transcription. *Adv. Genet.* **43**, 71–102 (2001).
- Boehmer, P. E. & Emmerson, P. T. *Escherichia coli* RecBCD enzyme: inducible overproduction and reconstitution of the ATP-dependent deoxyribonuclease from purified subunits. *Gene* **102**, 1–6 (1991).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank D. Julin for advice on protein purification; M. Dillingham, M. Spies and S. Kowalczykowski for sharing their unpublished information⁷; S. Amundsen for permission to cite unpublished results; J. Cooper, M. Gellert, N. Maizels, R. Strong and our colleagues for comments on the manuscript; and L. Caldwell and staff for help with the electron microscopy. This research was supported by grants from the NIH.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.R.S. (gsmith@fhcr.org).

RecBCD enzyme is a bipolar DNA helicase

Mark S. Dillingham[†]*, Maria Spies* & Stephen C. Kowalczykowski*

* Sections of Microbiology and of Molecular and Cellular Biology, Center for Genetics and Development, University of California, Davis, California 95616, USA

Escherichia coli RecBCD is a heterotrimeric helicase/nuclease that catalyses a complex reaction in which double-strand breaks in DNA are processed for repair by homologous recombination¹. For some time it has been clear that the RecB subunit possesses a 3' → 5' DNA helicase activity^{2–4}, which was thought to drive DNA translocation and unwinding in the RecBCD holoenzyme. Here we show that purified RecD protein is also a DNA helicase, but one that possesses a 5' → 3' polarity. We also show that the RecB and RecD helicases are both active in intact RecBCD, because the enzyme remains capable of processive DNA unwinding when either of these subunits is inactivated by mutation. These findings point to a bipolar translocation model for RecBCD in which the two DNA helicases are complementary, travelling with opposite polarities, but in the same direction, on each strand of the antiparallel DNA duplex. This bipolar motor organization helps to explain various biochemical properties of RecBCD, notably its exceptionally high speed and processivity, and offers a mechanistic insight into aspects of RecBCD function.

RecBCD enzyme processes DNA breaks for repair by homologous recombination by means of an elaborate reaction involving coordinated and regulated helicase and nuclease activities. After binding specifically to a double-stranded DNA end, this 330-kDa heterotrimer uses the free energy of ATP hydrolysis to translocate into and separate the duplex while preferentially degrading the 3'-terminated nascent single strand⁵. On encountering the recombination hotspot Chi, an octameric DNA sequence that is recognized as single-stranded DNA (ssDNA) by the enzyme approaching from its 3' side⁶, the frequency of cleavage is reduced and its polarity is switched to the 5'-terminated strand⁷. Because translocation and unwinding continue after recognition of Chi, the final product is a duplex DNA with a single-stranded DNA tail terminated at its 3' end with the Chi sequence. RecBCD is also capable of loading the RecA protein onto this ssDNA tail⁸ to form a substrate for DNA-strand invasion, the next step in the general pathway for

[†] Present address: National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK.

homologous recombination⁹. RecBCD is of particular interest because of its enigmatic biochemical properties. Isolated RecB protein contains helicase motifs and a carboxy-terminal nuclease domain, and displays modular 3' → 5' helicase and nuclease activities^{2,3,10}. It is related, and behaves similarly, to the Rep/UvrD/PcrA family of superfamily I (SF1) DNA helicases, which are poorly processive enzymes that initiate preferentially from a ssDNA region flanking the duplex^{2,3,10}. In contrast, RecBCD protein is among the most processive and rapid helicases characterized, unwinding 30,000 base pairs per binding event at 1,000 base pairs s⁻¹ (ref. 11), with a strong preference for initiation from a duplex end. The similarity between RecB and the Rep/UvrD/PcrA helicase family implies that the complex biochemistry of RecBCD can be attributed to additional (non-helicase) domains within RecB protein, or to the RecC and RecD subunits. Little is known about the isolated RecC protein, which is devoid of sequence motifs that could provide clues to its role, but the specificity of Chi sequence recognition is altered by mutations in this subunit¹. RecD contains SF1 helicase motifs and is a ssDNA-dependent ATPase¹² but, until now, helicase activity has not been demonstrated.

We and others^{12,13} have experienced difficulty in obtaining native RecD protein, and so a His-tagged RecD protein (^{his}RecD) was purified to near homogeneity by refolding on a Ni²⁺-affinity column (Fig. 1a). To serve as a control, a mutant ^{his}RecD protein (^{his}RecD^{K177Q}) with a Lys → Gln substitution in the Walker A box (helicase motif I) was also purified. Mutation of the Walker A box lysine drastically decreases ATP hydrolysis in many helicases,

including the isolated RecB and RecD subunits^{12,14}. With the use of a coupled ATP hydrolysis assay, ^{his}RecD was shown to be an ATPase whose activity is stimulated 70-fold by ssDNA: $K_m(\text{poly}(\text{dT})) = 9 \mu\text{M}$ nucleotides; $V_{\text{max}} = 5 \text{ mol ATP per mol } ^{\text{his}}\text{RecD per second}$, which represents a minimum limit on the turnover number (k_{cat}) if 100% of the protein is active (Fig. 1b). Linear blunt-ended duplex DNA did not stimulate the ATPase activity (data not shown). To address the question of whether ^{his}RecD protein possesses DNA helicase activity, we tested its ability to separate duplex DNA substrates formed from short oligonucleotides (Table 1). ^{his}RecD was capable of separating the 5' → 3' substrate: a 40-base-pair duplex flanked by a 5'-terminated ssDNA tail (Fig. 2). Helicase activity was undetectable on the equivalent blunt duplex and was decreased at least 10-fold on the 3' → 5' substrate: a 40-base-pair duplex flanked by a 3'-terminated ssDNA tail (Fig. 2a, b). Unwinding of the 5' → 3' substrate was absolutely dependent on ^{his}RecD and ATP hydrolysis (Fig. 2c). Removal of ATP or inclusion of EDTA resulted in no strand displacement activity, which is consistent with the expected requirement for a Mg²⁺-dependent ATPase activity to support the helicase reaction. The non-hydrolysable ATP analogues

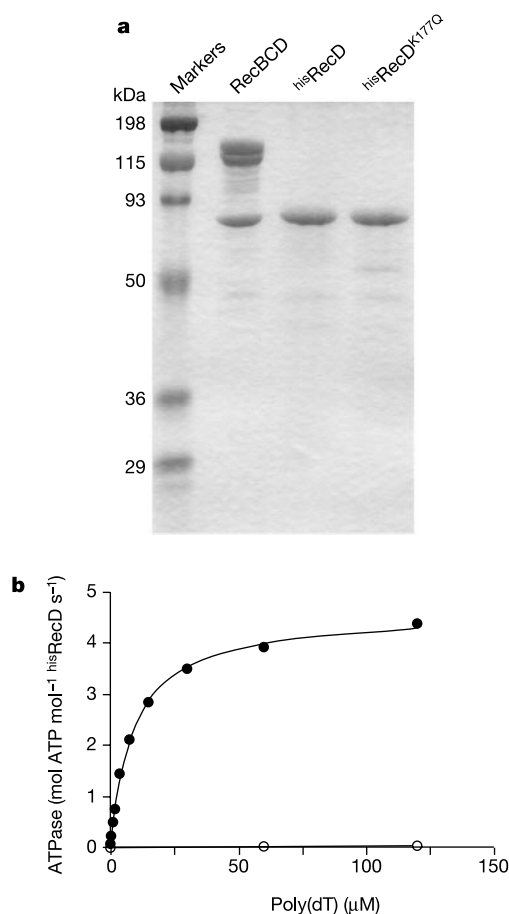


Figure 1 Purified ^{his}RecD protein is a ssDNA-dependent ATPase. **a**, SDS-PAGE gel (12% polyacrylamide) showing molecular mass markers, 7.5 pmol RecBCD, 15 pmol ^{his}RecD and 15 pmol ^{his}RecD^{K177Q}. **b**, Dependence of the ATPase activities of ^{his}RecD (filled circles) and ^{his}RecD^{K177Q} (open circles) on ssDNA.

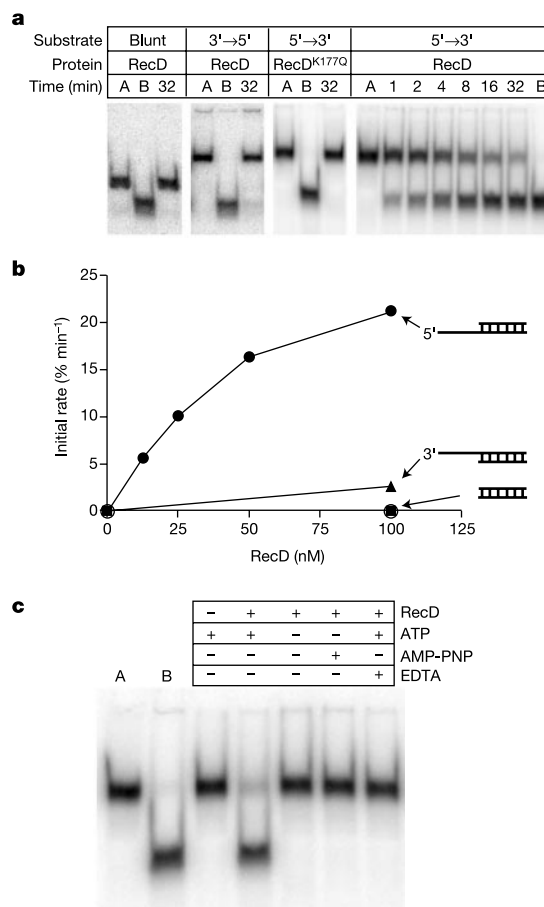


Figure 2 ^{his}RecD is a 5' → 3' DNA helicase. **a**, Helicase assays in which 100 nM ^{his}RecD or ^{his}RecD^{K177Q} was incubated with 1 nM DNA substrate and 2.5 mM ATP for the times indicated. Lanes A and B are controls for annealed and heat-denatured substrate respectively. Experiments were performed with three different DNA substrates (Table 1) and either wild-type or K177Q mutant ^{his}RecD. **b**, Rates of helicase activity, determined from the initial slopes of helicase time-course assays, for blunt (filled squares), 3' → 5' (filled triangles), and 5' → 3' (filled circles) DNA substrates using ^{his}RecD. Unwinding of the 5' → 3' substrate by ^{his}RecD^{K177Q} (open circles) is very poor. **c**, Controls demonstrating that ATP hydrolysis is required for ^{his}RecD-catalysed helicase activity on the 5' → 3' DNA substrate. Helicase assays were performed for 30 min with the indicated components.

Table 1 DNA substrates used for strand displacement helicase assays

Name	Substrate structure
Blunt	3' - GTTATGCGTTTGGCGGAGAGGGGCGCGCAACCGGCTAAGT - 5' 5' - *CAATACGCAAACCGCCTCTCCCCGCGGTTGGCCGATTCA - 3'
3' → 5'	3' - CTCGCGTCGCTCAGTCACTCGCTCCCTTCGCGGTTATGCGTTTGGCGGAGAGGGGCGCGCAACCGGCTAAGT - 5' 5' - *CAATACGCAAACCGCCTCTCCCCGCGGTTGGCCGATTCA - 3'
5' → 3'	3' - GTTATGCGTTTGGCGGAGAGGGGCGCGCAACCGGCTAAGT* - 5' 5' - GAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGGTTGGCCGATTCA - 3'

The asterisk indicates the position of the ³²P-phosphate radiolabel.

AMP-PNP (Fig. 2c) and ATP-γS (data not shown) did not support helicase activity but did bind to ^{his}RecD as shown by virtue of their ability to inhibit the ATPase activity (Supplementary Information). The ATPase¹² and 5' → 3' helicase activities are drastically decreased in the ^{his}RecD^{K177Q} mutant protein (Figs 1b and 2), which indicates that they are intrinsic to the wild-type RecD polypeptide and dependent on the Walker A box lysine residue.

Because protein prepared by refolding might be inactive, we sought to detect a functional interaction with RecBC by assaying for the production of Chi-specific ssDNA fragments, a biochemical property associated exclusively with the RecBCD holoenzyme (Fig. 3a). As expected, RecBCD enzyme produces Chi-specific fragments and degrades the DNA to an extent typically observed⁵, whereas RecBC enzyme (which is largely devoid of nuclease activity) produces very few Chi-specific fragments but is active as a helicase, generating full-length ssDNA product (Fig. 3b). ^{his}RecD protein has no activity on this plasmid-length duplex DNA substrate. However, when RecBC and ^{his}RecD were mixed, a substantial increase in Chi-specific fragments was observed, indicating the reconstitution of

functional RecBCD enzyme. Note that ^{his}RecD protein stimulates the rate of substrate use (that is, the helicase activity) by RecBC. This effect was also observed when a dye displacement helicase assay was used (data not shown).

To address the question of whether the RecB and RecD helicases are active within the RecBCD holoenzyme, we purified RecBCD enzymes containing Lys → Gln mutations in the Walker A box (helicase motif I) of either RecB (K29Q) or RecD (K177Q), and examined their DNA-unwinding capacities. As expected¹, wild-type RecBCD rapidly unwound a linearized plasmid DNA lacking Chi sequences, producing full-length ssDNA product and substantially degrading the DNA by means of its nuclease activity (Fig. 4). Remarkably, both the RecB^{K29Q}CD and the RecBCD^{K177Q} enzymes retained potent helicase activity, showing that either motor can support processive DNA translocation, and indicating strongly that RecB and RecD are simultaneously active in the wild-type holoenzyme. As expected, and in contrast to these single-mutant proteins, the RecB^{K29Q}CD^{K177Q} double-mutant protein is inactive both as a ssDNA-dependent ATPase¹⁴ and as a helicase on long duplex DNA substrates¹⁵.

In this study we have demonstrated that RecD protein is a 5' → 3' DNA helicase. The most closely related proteins to RecD include the TraI and Dda SF1 helicases, both of which also possess 5' → 3' DNA helicase activity^{16,17}. Similarly, RecB protein is related to the Rep/UvrD/PcrA class of SF1 helicases, which all possess 3' → 5' polarity. It is well established that RecB possesses a 3' → 5' DNA helicase activity, as this has been demonstrated for both RecB^{2,3}, (M.S.D. and S.C.K., unpublished observations) and RecBC⁴ proteins. We have

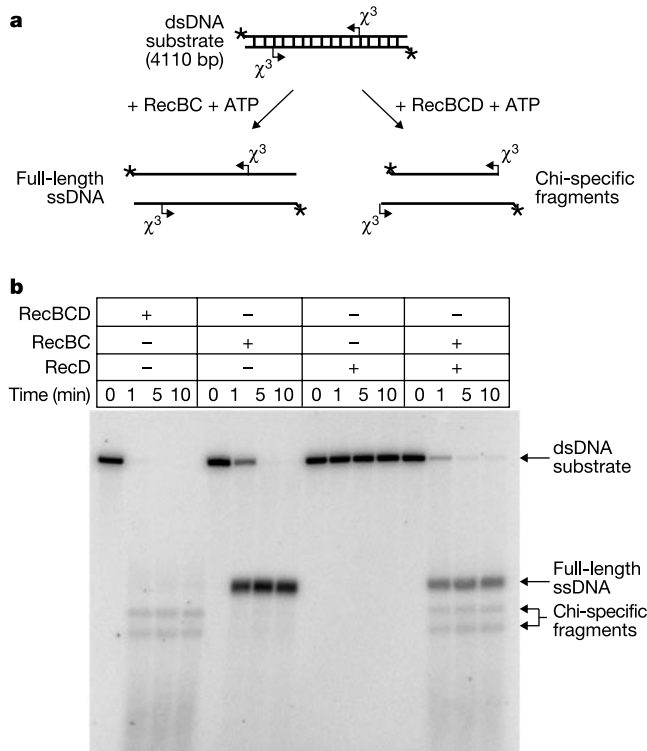


Figure 3 Reconstitution of the Chi-specific cleavage activity of RecBCD from purified RecBC and ^{his}RecD proteins. **a**, A 5'-radiolabelled (asterisks) linearized plasmid containing 'triple Chi' (χ³) sequences is unwound to full-length ssDNA by RecBC, whereas RecBCD enzyme processes this substrate to Chi-terminated ssDNA fragments. dsDNA, double-stranded DNA. **b**, The substrate was incubated with the enzyme(s) indicated in the presence of ATP as described in Methods.

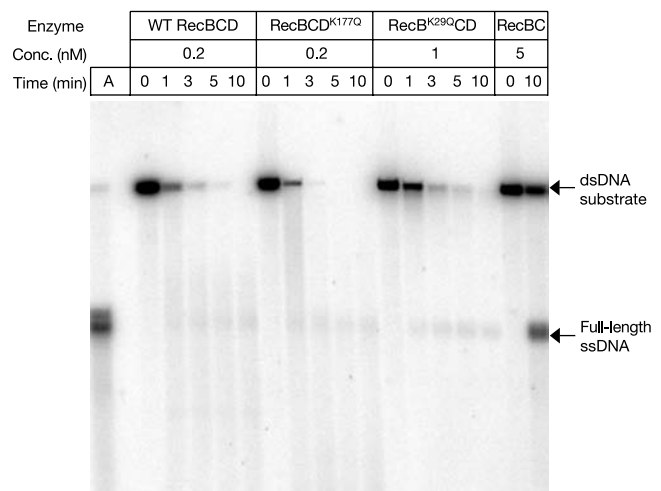


Figure 4 The RecB and RecD helicase subunits are both active in the RecBCD holoenzyme. Unwinding of a 5'-radiolabelled linearized plasmid (lacking Chi sequences) by wild-type RecBCD, wild-type RecBC and mutant holoenzymes containing a substitution in the Walker A motif of either the RecB or RecD subunit (see the text). Lane A contained a heat-denatured sample of the substrate to mark the position of full-length ssDNA. The DNA substrate was incubated with the enzyme indicated in the presence of ATP as described in Methods. dsDNA, double-stranded DNA; wt, wild-type.

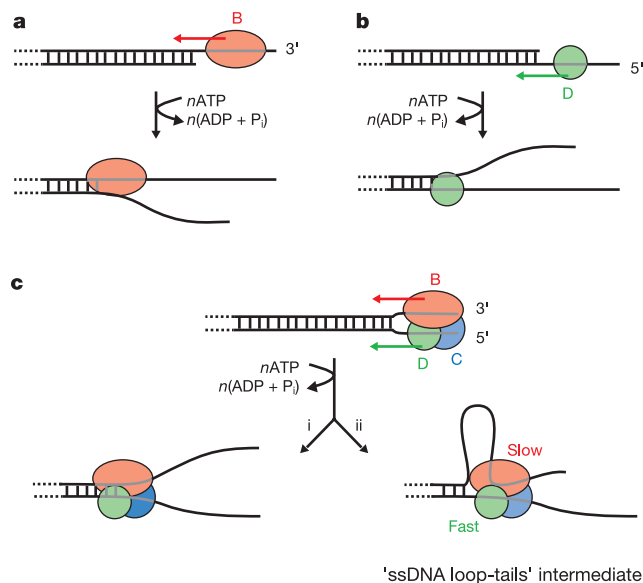


Figure 5 A bipolar DNA helicase translocation model. **a**, RecB is a 3' → 5' helicase; **b**, RecD is a 5' → 3' DNA helicase. Experiments with other helicases (see the text) indicate that helicase polarity relates to a unidirectional ssDNA translocation activity as illustrated. **c**, Because RecB and RecD helicases are of opposite polarity relative to ssDNA, they complement the antiparallel DNA structure, producing cooperative movement that is unidirectional relative to the duplex. Before translocation, the RecB and RecD proteins are positioned on the 3' - and 5' -terminated single strands respectively, in accordance with crosslinking data²³ and their observed polarities. The position of the RecC protein is arbitrary. Pathway i: on ATP hydrolysis the two helicases move uniformly along the DNA, unwinding the duplex as they progress. Pathway ii: alternatively, if the motors move independently and at unequal speeds, then a bipolar translocation mechanism would explain the generation of the 'ssDNA loop-tails' unwinding intermediates that have been observed by electron microscopy²⁵ (see the text). The relative rates of the RecB and RecD helicases are not determined in this work. This assignment of fast and slow activities is based on ref. 15, in which it was shown that RecD subunit is the faster motor under the conditions used.

further shown that both the RecB and RecD subunits are capable of driving the unwinding of DNA in the RecBCD holoenzyme. Why might RecBCD enzyme possess two DNA helicase subunits of opposite polarity? Crystallographic¹⁸ and biochemical^{19–21} data show that the SF1 helicases PcrA and Dda contain an autonomous unidirectional ssDNA translocation activity that is consistent with the polarity of their helicase activities. The structural work suggests that this ssDNA translocation motor is formed by the 'helicase' motifs. Consequently, SF1 helicase motifs are suggested to be a blueprint for a ssDNA motor: a modular structure capable of providing the translocation function in a DNA helicase or other DNA-processing protein²². At first glance, the observation of two helicase activities of opposite polarity contained within a single protein complex might seem unusual. However, the notion that the two proteins function as ssDNA motors suggests an elegant mechanism by which they might cooperate to translocate along a DNA duplex. The antiparallel nature of the duplex would allow the two proteins to bind to opposite strands at the DNA break, yet to translocate in the same direction relative to the duplex (Fig. 5). This simple bipolar helicase model helps to explain many of the enigmatic properties of RecBCD enzyme. First, the blunt-end loading of the enzyme might reflect the positioning of two ssDNA motors side by side on each strand of the duplex. This suggestion is consistent with crosslinking data that place the RecB and RecD subunits on the 3' - and 5' -terminated strands, respectively²³. Second, a bipolar helicase organization can explain the exceptional processivity of

RecBCD: the two motors would have to dissociate simultaneously to release the holoenzyme from its DNA track. Indeed, RecBCD is more processive and faster than RecBC enzyme^{11,24} (Fig. 3), which might indicate that the RecD motor is faster than RecB, or that RecD augments that activity of RecB. Last, our model can explain the 'ssDNA loop-tails' unwinding intermediates observed by electron microscopy²⁵. If two linked motors were to travel on opposite strands at unequal speeds, the faster motor would be associated with a long ssDNA tail, whereas the slower motor would be associated with a ssDNA loop ahead of its itself and a trailing ssDNA tail (Fig. 5c, pathway ii). In this model, the leading motor is a genuine helicase whereas the slower, second motor acts simply as a ssDNA translocase. However, our data with the mutant holoenzymes (Fig. 4) suggest that either motor subunit can function as the leading helicase, which is consistent with general models for helicase activity in which a single ssDNA motor is the vital component. Recent electron microscopy data¹⁵ support these ideas, showing that mutation of helicase motif I in either the RecB or the RecD subunit alters the looped intermediates formed by RecBCD in a manner that suggests a 'dual-helicase' model. Our proposal of a bipolar motor organization raises several questions about the RecBCD mechanism. Can RecBCD bypass DNA damage or gaps in both strands as it progresses along the duplex⁴? Importantly, because the RecD subunit is required for the production of Chi-specific fragments by RecBCD (Fig. 3), and given the prevailing models of RecD subunit inactivation in response to recombination hotspot recognition¹, does this translocation mechanism provide an opportunity for the differential DNA processing observed before and after the recognition of Chi? Last, it will be of interest to see how bipolar helicase motors might be employed in other macromolecular assemblies that process DNA structure, such as the TFIIH complex: a basal transcription and nucleotide excision repair factor containing the 3' → 5' XPB and 5' → 3' XPD helicases²⁶. □

Methods

Protein expression and purification

^{his}RecD and ^{his}RecD^{K177Q} proteins were expressed in the form of inclusion bodies by using the pET15b vector system (Novagen), which introduces an amino-terminal histidine tag on a 20-amino-acid leader sequence. Purification was performed under denaturing conditions with Ni²⁺-chelating Sepharose resin, followed by an on-column refolding technique (Supplementary Information). The integrity of the cloned *recD* and *recD*^{K177Q} genes was confirmed by DNA sequencing (at the DBS sequencing facility, University of California, Davis, California). The *recD* nucleotide sequence was identical to that found in the *E. coli* K12 complete genome (GenBank accession number NC_000913). Because of the nature of the method used to purify the RecD protein, it is possible that the preparation was not fully active. RecBC protein was expressed with the pPB520 and pPB700 plasmids¹³ in a Δ *recBCD* background (V330) containing a plasmid expressing LacI⁴, and purified with a RecBCD protocol as described⁶. RecBCD protein was expressed and purified as described⁶. Expression strains for the RecB^{K29Q}CD and RecBCD^{K177Q} proteins^{27,28} in a Δ *recBCD* background were a gift from D. A. Julin (University of Maryland, College Park, Maryland, USA). The mutant proteins were expressed as described^{27,28} and purified with the protocol for the wild-type enzyme⁶.

ATP hydrolysis assays

ATP hydrolysis was measured spectrophotometrically by coupling it to NADH oxidation²⁹. Assays were performed with 80 nM ^{his}RecD or 240 nM ^{his}RecD^{K177Q} at room temperature (22 °C) in a buffer containing 20 mM Tris-HCl pH 7.5, 50 mM NaCl, 3 mM MgCl₂, 4 mM dithiothreitol, 35 U ml⁻¹ pyruvate kinase, 20 U ml⁻¹ lactate dehydrogenase, 2 mM phosphoenolpyruvate, 0.08 mg ml⁻¹ NADH, 2 mM (saturating) ATP, and the indicated concentration (in nucleotides) of poly(dT) ssDNA. Data were fitted directly to the Michaelis-Menten equation with the program GraphPad Prism 3.02 (GraphPad Software).

Strand-displacement DNA helicase assays

³²P-labelled helicase substrates were prepared by annealing the oligonucleotides shown in Table 1, which had been purified by polyacrylamide-gel electrophoresis. Labelled oligonucleotide (3 pmol) was annealed to its partner (5 pmol) to form the final substrate, which was purified with a G25 column (AP-Biotech). Assays were performed at room temperature in a buffer containing 20 mM Tris-HCl pH 7.5, 17.5 mM NaCl, 2 mM MgCl₂, 4 mM dithiothreitol, 2.5 mM ATP and 1 nM (in molecules) DNA substrate. Where indicated, 2.5 mM AMP-PNP or 10 mM EDTA was included in the reaction. Reactions were initiated with RecD (100 nM or as indicated) and terminated with 0.4% w/v SDS, 40 mM EDTA, 8% v/v glycerol, 0.1% w/v bromophenol blue and 200 nM unlabelled

oligonucleotide as a trap to prevent the reannealing of the labelled strand. However, the presence of trap DNA in the stop buffer made no significant difference to our results. The oligonucleotide reannealing rate was shown empirically to be insignificant over the time course of our assays. Reaction products were separated on native 15% polyacrylamide gels, which were dried on DEAE paper and quantified with a STORM PhosphorImager and ImageQuant software (Molecular Dynamics).

Plasmid DNA unwinding and Chi-specific fragment production assays

Assays were performed with slight modifications to a previously described method³⁰. *NdeI*-linearized, ³²P-labelled pBR322 χ^{3F3H} plasmid containing two sets of 'triple-Chi' sequences was used as the substrate in reactions (initiated with 2 mM ATP after a 2-min preincubation of all other components) and performed at room temperature with either 0.5 nM RecBCD, 5 nM RecBC or 50 nM ^{his}RecD protein, as indicated, in a buffer containing 25 mM Tris-acetate pH 7.5, 6 mM magnesium acetate, 1 mM dithiothreitol, 20 μ M (nucleotides) DNA and 2 μ M *E. coli* single-stranded DNA-binding protein. Assays with DNA lacking Chi sequences were performed with the same method, but with *NdeI*-linearized, ³²P-labelled pBR322 plasmid as the substrate, and using the enzyme concentrations indicated.

Received 7 November 2002; accepted 8 April 2003; doi:10.1038/nature01673.

- Arnold, D. A. & Kowalczykowski, S. C. In *Encyclopedia of Life Sciences* <http://www.els.net> (Nature Publishing Group, London, 1999).
- Phillips, R. J., Hickleton, D. C., Boehmer, P. E. & Emmerson, P. T. The RecB protein of *Escherichia coli* translocates along single-stranded DNA in the 3' to 5' direction: a proposed ratchet mechanism. *Mol. Gen. Genet.* **254**, 319–329 (1997).
- Yu, M., Souaya, J. & Julin, D. A. The 30-kDa C-terminal domain of the RecB protein is critical for the nuclease activity, but not the helicase activity, of the RecBCD enzyme from *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **95**, 981–986 (1998).
- Bianco, P. R. & Kowalczykowski, S. C. Step size measurements on the translocation mechanism of the RecBC DNA helicase. *Nature* **405**, 368–372 (2000).
- Dixon, D. A. & Kowalczykowski, S. C. The recombination hotspot Chi is a regulatory sequence that acts by attenuating the nuclease activity of the *E. coli* RecBCD enzyme. *Cell* **73**, 87–96 (1993).
- Bianco, P. R. & Kowalczykowski, S. C. The recombination hotspot Chi is recognized by the translocating RecBCD enzyme as the single strand of DNA containing the sequence 5'-GCTGGTGG-3'. *Proc. Natl Acad. Sci. USA* **94**, 6706–6711 (1997).
- Anderson, D. G. & Kowalczykowski, S. C. The recombination hot spot Chi is a regulatory element that switches the polarity of DNA degradation by the RecBCD enzyme. *Genes Dev.* **11**, 571–581 (1997).
- Anderson, D. G. & Kowalczykowski, S. C. The translocating RecBCD enzyme stimulates recombination by directing RecA protein onto ssDNA in a Chi-regulated manner. *Cell* **90**, 77–86 (1997).
- Kowalczykowski, S. C., Dixon, D. A., Eggleston, A. K., Lauder, S. D. & Rehrauer, W. M. Biochemistry of homologous recombination in *Escherichia coli*. *Microbiol. Rev.* **58**, 401–465 (1994).
- Zhang, X. J. & Julin, D. A. Isolation and characterization of the C-terminal nuclease domain from the RecB protein of *Escherichia coli*. *Nucleic Acids Res.* **27**, 4200–4207 (1999).
- Roman, L. J., Eggleston, A. K. & Kowalczykowski, S. C. Processivity of the DNA helicase activity of *Escherichia coli* RecBCD enzyme. *J. Biol. Chem.* **267**, 4207–4214 (1992).
- Chen, H. W., Ruan, B., Yu, M., Wang, J. & Julin, D. A. The RecD subunit of the RecBCD enzyme from *Escherichia coli* is a single-stranded DNA-dependent ATPase. *J. Biol. Chem.* **272**, 10072–10079 (1997).
- Boehmer, P. E. & Emmerson, P. T. *Escherichia coli* RecBCD enzyme: inducible overproduction and reconstitution of the ATP-dependent deoxyribonuclease from purified subunits. *Gene* **102**, 1–6 (1991).
- Chen, H.-W., Randle, D. E., Gabbidon, M. & Julin, D. A. Functions of the ATP hydrolysis subunits (RecB and RecD) in the nuclease reactions catalyzed by the RecBCD enzyme from *Escherichia coli*. *J. Mol. Biol.* **278**, 89–104 (1998).
- Taylor, A. F. & Smith, G. R. RecBCD enzyme is a DNA helicase with fast and slow motors of opposite polarity. *Nature* **XXX**, XXX–XXX (2003).
- Lahue, E. E. & Matson, S. W. *Escherichia coli* DNA helicase I catalyzes a unidirectional and highly processive unwinding reaction. *J. Biol. Chem.* **263**, 3208–3215 (1998).
- Jongeneel, C. V., Formosa, T. & Alberts, B. M. Purification and characterization of the bacteriophage T4 dda protein. A DNA helicase that associates with the viral helix-destabilizing protein. *J. Biol. Chem.* **259**, 12925–12932 (1984).
- Velankar, S. S., Soutanas, P., Dillingham, M. S., Subramanya, H. S. & Wigley, D. B. Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. *Cell* **97**, 75–84 (1999).
- Morris, P. D. & Raney, K. D. DNA helicases displace streptavidin from biotin-labeled oligonucleotides. *Biochemistry* **38**, 5164–5171 (1999).
- Dillingham, M. S., Wigley, D. B. & Webb, M. R. Demonstration of unidirectional single-stranded DNA translocation by PcrA helicase: measurement of step size and translocation speed. *Biochemistry* **39**, 205–212 (2000).
- Dillingham, M. S., Wigley, D. B. & Webb, M. R. Direct measurement of single-stranded DNA translocation by PcrA helicase using the fluorescent base analogue 2-aminopurine. *Biochemistry* **41**, 643–651 (2002).
- Singleton, M. R. & Wigley, D. B. Modularity and specialization in superfamily 1 and 2 helicases. *J. Bacteriology* **184**, 1819–1826 (2002).
- Ganesan, S. & Smith, G. R. Strand-specific binding to duplex DNA ends by the subunits of the *Escherichia coli* RecBCD enzyme. *J. Mol. Biol.* **229**, 67–78 (1993).
- Koranyi, F. & Julin, D. A. Kinetics and processivity of ATP hydrolysis and DNA unwinding by the RecBC enzyme from *Escherichia coli*. *Biochemistry* **32**, 4873–4880 (1993).
- Taylor, A. F. & Smith, G. R. Unwinding and rewinding of DNA by the RecBC enzyme. *Cell* **22**, 447–457 (1980).
- van Brabant, A. J., Stan, R. & Ellis, N. A. DNA helicases, genomic instability, and human genetic disease. *Annu. Rev. Genom. Hum. Genet.* **1**, 409–459 (2000).
- Koranyi, F. & Julin, D. A. Alteration by site-directed mutagenesis of the conserved lysine residue in the ATP-binding consensus sequence of the RecD subunit of the *Escherichia coli* RecBCD enzyme. *J. Biol. Chem.* **267**, 1727–1732 (1991).
- Hsieh, S. & Julin, D. A. Alteration by site-directed mutagenesis of the conserved lysine residue in the consensus ATP-binding sequence of the RecB protein of *Escherichia coli*. *Nucleic Acids Res.* **20**, 5647–5653 (1992).
- Kreuzer, K. N. & Jongeneel, C. V. *Escherichia coli* phage T4 topoisomerase. *Methods Enzymol.* **100**, 144–160 (1983).
- Arnold, D. A., Bianco, P. R. & Kowalczykowski, S. C. The reduced levels of Chi recognition exhibited by the RecBC^{1004D} enzyme reflect its recombination defect *in vivo*. *J. Biol. Chem.* **273**, 16476–16486 (1998).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank N. Handa and K. Morimatsu for preparative work; D. Julin for his generosity; A. Taylor and G. Smith for discussions and for sharing their unpublished data¹⁵; and the members of the Kowalczykowski laboratory for their critical reading of the manuscript. This work was supported by a Wellcome Trust Travelling Research Fellowship to M.S.D., an American Cancer Society Postdoctoral Fellowship to M.S. and by an NIH grant to S.C.K.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.C.K. (skowalczykowski@ucdavis.edu).

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Senior European Sales Manager:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

Netherlands/ Italy/

Iberia/ Belgium:

Evelina Rubio Hakansson (4973)

Scandinavia: Silje Opstrup (4994)

France/ Switzerland:

Amelie Pequignot (4974)

Production Manager: Billie Franklin

To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Hind Berrada (4935)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Office

Germany/ Austria:

Patrick Phelan, Odo Wulffen

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mail: p.phelan@nature.com

o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

US Advertising Coordinator:

Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Haraikatomachi,
Shinjuku-ku,
Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Hideki Watanabe

e-mail: h.watanabe@naturejpn.com

naturejobs

Career-change carousel

The traditional internship — a short-term placement with little or no pay — is often associated with undergraduates. Once they have gained their degree, this experience will often get them a place on a low rung of the career ladder.

There are signs that science PhDs seeking alternative careers are now adapting that strategy to their own ends. They are taking intermediate-level jobs in a different sector before taking the big leap into another career that involves considerable commitment or retraining.

Kelvin Stott, founder and chief scientific officer of Manchester-based biotechnology company Senexis, worked for London-based management consultants McKinsey to get the business skills he needed to launch his own company. While doing a postdoc at Cambridge a few years ago, Stott developed a technology he thought had commercial potential. But he had only vague notions of how to obtain seed funding and become a business. At McKinsey he gained management insight and ideas about how to raise capital.

This tactic isn't limited to would-be biotech entrepreneurs. Rather than do a postdoc after her PhD, Hayley French pursued a master's degree in intellectual-property law and worked in that field for a time before taking the even bigger step of training to become a fully fledged lawyer (see *Naturejobs* **423**, 666–667; 2003). Nor is it just a way into the private sector. Tshaka Cunningham, a graduate student at the Rockefeller University in New York, worked for a few years at Bristol-Myers Squibb before returning to academia (*Naturejobs* **418**, 3; 2002).

With academic jobs in limited supply, it is not surprising if prudent scientists make similar forays into other sectors instead of a second postdoc, or even in place of a first.

Paul Smaglik
Naturejobs editor



Contents

POSTDOCS AND STUDENTS

German postdocs face
a difficult future

p900

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

Home truths

Economic conditions are making it harder for young German postdocs to stay in their home country. But a few programmes offer some hope, says Quirin Schiermeier.

The prospect of sunny weather and year-round beach life was not why Giovanni Galizia moved last year from Germany to the University of California, Riverside. The 40-year-old neurobiologist, who studies the insect olfactory system, can already look back on more than ten successful years in research. But he started to feel increasingly ill at ease with the German research system, which, he says, fails to offer a clear career perspective — even to postdoctoral scientists with proven talent. So when offered a tenure-track position as associate professor at Riverside, he lost little time in accepting. The Free University in Berlin would have gladly retained him, but could not offer him such attractive conditions.

Galizia is not alone. A combination of changes in research policy and economic problems have made many young German researchers consider either leaving the country or abandoning the bench. A university law introduced last year allows German scientists to spend no more than 12 years on fixed-term contracts, including PhDs and postdoctoral positions, before they obtain a tenured university post (see *Nature* **415**, 257–258; 2002), and this is driving some abroad. The economic crisis, which has reduced science funding, despite recent promises of increases, is forcing others to leave (see *Nature* **420**, 452; 2002).

Giovanni Galizia didn't move to California just for the beach.

As a result of these conditions, German universities are finding it increasingly hard to attract the best young scientists, says Jürgen Mlynek, the rector of the Humboldt University in Berlin. Because there are too few tenure-track positions, many postdocs prefer a more secure position abroad or in industry, he says.

FIGHTING BACK

But Mlynek points to a number of promising new initiatives to enhance the status of young scientists at Germany's 100 or so research universities and numerous

non-university institutes. Junior professorships, for example, were introduced last year to give postdoctoral researchers full scientific independence at an earlier age. And the main granting agencies, such as the publicly funded DFG and the private Volkswagen Foundation, each operate tailor-made programmes for top young researchers.

The DFG's prestigious Emmy Noether programme, for example, includes repatriation grants for young German scientists abroad. The Volkswagen Foundation's new Tandem Programme targets pairs of young postdocs with interdisciplinary ideas in non-established combinations of fields. And the foundation's Lichtenberg professorships offer researchers with two or three years' postdoctoral experience tenure-track positions in a wide range of innovative fields.

In general, the competition for grants is getting

Washington outpost

Last year the German research foundation (DFG) opened a liaison office in Washington, DC. Its goals are threefold: to help German scientists who have moved to the United States to find funding; to make US scientists more aware of German research activity; and to help German scientists

move back to Germany once opportunities arise. The Washington office aims to establish a network of German scientists in the United States who have previously received DFG funding, by matching up researchers funded from various public and non-profit sources **Paul Smaglik** DFG Washington office www.dfg-usa.org



POSTDOCS & STUDENTS

"I feel mucked around. I had apparently taken part in a race without knowing that I was in it," she says.

After several applications for jobs in the pharmaceutical industry remained unanswered, and frustrated by the pointlessness of carrying on, she gave it all up. She is now preparing for a second career in science journalism or in an editorial office.

"This is a typical example of how Germany tends to over-regulate people's lives," says Axel Meyer, head of evolutionary biology at the University of Konstanz, one of the rising stars among the younger generation of German life scientists.

Experienced scientists such as Kortenjann, who for whatever reason have not obtained a tenured position, can still be very valuable, says Meyer. He recently also lost a German postdoc who, facing the approaching 12-year limit, left for Canada. "I would have liked him to stay and I had funding for him, but the new rule required that he had to leave Germany's university system," Meyer says.

SIGNS OF HOPE

By and large, however, Germany remains an attractive place for young scientists, says Gregor Markl, a mineralogist at the University of Tübingen, who four years ago became the youngest full professor in Germany at the age of 28. He admits that it is getting harder in Germany to succeed in research. "But let's not decry this country," he says. "The actual opportunities here are better than the general mood." Markl stresses the importance of helping young scientists to make the right decisions before research careers end up in a dead-end street.

"Of course it is not easy to tell somebody that his or her gifts may not be sufficient, but I wouldn't hesitate to tell someone mediocre that it may be better for him or her to look out for something else," he says.

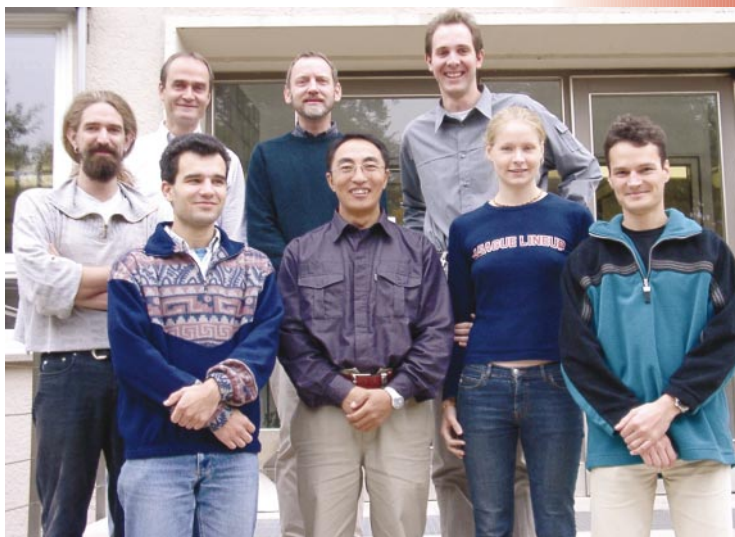
But what will a 32-year-old professor advise a good postdoc, who may well be older than himself? "That hard work alone is not enough," says Markl: "It also takes a lot of talent, imaginative power, a real urge to do research — and a good bit of fortune."

Quirin Schiermeier is *Nature's* German correspondent.

♦ www.volkswagen-stiftung.de/english.html

♦ www.dfg.de/english/index_en.html

Gregor Markl (back row, far right) thinks Germany is still a good place for young researchers.



Not all doom and gloom: Wolf Schmidt supervises an experiment in Jena.

harder. The overall rate of successful applications at the Volkswagen Foundation has dropped from almost 60% to 44% since 1995, and in the wake of the current budget constraints, many DFG-funded projects are being downsized.

UNKINDEST CUTS

Many young researchers can tell tales of unexpected cuts. Wolf Schmidt, a 35-year-old theoretical physicist at the University of Jena, recently received a letter from the university's chancellor telling him that his group's budget for small equipment and travelling needs to be reduced by €3,500 (US\$4,100). "It doesn't sound a lot," he says. "But it is very annoying to have two computers less at times when we have more graduate students than ever before."

The difficulties are more noticeable in Germany's poorer *Länder* (states), in the east and north, than in the rich south, where universities continue to receive more generous support from state governments.

Earlier this year in Berlin, Mlynek announced an end to all new appointments at the Humboldt University if the government doesn't give it any more money (see *Nature* **423**, 101; 2003). As a result of this cash crunch, young researchers there describe the atmosphere at the brink of bankruptcy as paralyzing.

"Berlin is a crazy city, but I hope not so crazy as to intentionally rid itself of the best young scientists," says Ingo Bachmann, a cell biologist and neurobiologist who was appointed last year as a junior professor at the Charité, the Humboldt University's medical faculty. "I don't worry about my personal future," he adds, "but if Berlin is not able to pay for quality, then I will go somewhere else."

But many postdocs in Germany do have to worry about their future. Monika Kortenjann, for example, a biologist formerly at the Max Planck Institute of Immunobiology in Freiburg, is looking back at the age of nearly 40 on the ruins of an 11-year career in research. A fixed-term group leader at the institute since 1998, her contract was terminated last year, after the DFG cancelled two projects of a large collaborative grant in which she participated.

Kortenjann applied for a position at the University of Essen — only to find that her 'time' under the new university law had run out. She was flabbergasted to learn that the new rule applied to her, although her career had started many years before its introduction.